Binary and Multi-Bit Coding for Stable Random Projections

Ping Li

Department of Statistics and Biostatistics
Department of Computer Science
Rutgers University
Piscataway, NJ 08854, USA
pingli@stat.rutgers.edu

Abstract

We develop efficient binary (i.e., 1-bit) and multi-bit coding schemes for estimating the scale parameter of α-stable distributions. The work is motivated by the recent work on **one scan 1-bit compressed sensing** (sparse signal recovery) [12] using α-stable random projections, which requires estimating of the scale parameter at bits-level. Our technique can be naturally applied to data stream computations for estimating the α-th frequency moment. In fact, the method applies to the general scale family of distributions, not limited to α-stable distributions.

Due to the heavy-tailed nature of α-stable distributions, using traditional estimators will potentially need many bits to store each measurement in order to ensure sufficient accuracy. Interestingly, our paper demonstrates that, using a simple closed-form estimator with merely 1-bit information does not result in a significant loss of accuracy if the parameter is chosen appropriately. For example, when $\alpha = 0+$, 1, and 2, the coefficients of the optimal estimation variances using full (i.e., infinite-bit) information are 1, 2, and 2, respectively. With the 1-bit scheme and appropriately chosen parameters, the corresponding variance coefficients are 1.544, $\pi^2/4$, and 3.066, respectively. Theoretical tail bounds are also provided. Using 2 or more bits per measurements reduces the estimation variance and importantly, stabilizes the estimate so that the variance is not sensitive to parameters. With look-up tables, the computational cost is minimal.

Extensive simulations are conducted to verify the theoretical results. The estimation procedure is integrated into the sparse recovery with one scan 1-bit compressed sensing. One interesting observation is that the classical "Bartlett correction" (for MLE bias correction) appears particularly effective for our problem when the sample size (number of measurements) is small.

1 Introduction

The research problem of interest is about efficient estimation of the scale parameter of the α-stable distribution using binary (i.e., 1-bit) and multi-bit coding of the samples. That is, given n i.i.d. samples,

$$y_j \sim S(\alpha, \Lambda_\alpha), \qquad j = 1, 2, ..., n \qquad (1)$$

from an α-stable distribution $S(\alpha, \Lambda_\alpha)$, we hope to estimate the scale parameter Λ_α by using only 1-bit or multi-bit information of $|y_j|$. Here we adopt the parameterization [22, 19] such that, if

$y \sim S(\alpha, \Lambda_\alpha)$, then the characteristic function is $E\left(e^{\sqrt{-1}yt}\right) = e^{-\Lambda_\alpha |t|^\alpha}$. Note that, under this parameterization, when $\alpha = 2$, $S(2, \Lambda_2)$ is equivalent to a Gaussian distribution $N(0, \sigma^2 = 2\Lambda_2)$. When $\alpha = 1$, $S(1, 1)$ is the standard Cauchy distribution.

1.1 Sampling from α-stable Distribution

Although in general there is no closed-form density of $S(\alpha, 1)$, we can sample from the distribution using a standard procedure provided by [5]. That is, one can first sample an exponential $w \sim exp(1)$ and a uninform $u \sim unif(-\pi/2, \pi/2)$, and then compute

$$s_\alpha = \frac{\sin(\alpha u)}{(\cos u)^{1/\alpha}}\left[\frac{\cos(u - \alpha u)}{w}\right]^{(1-\alpha)/\alpha} \sim S(\alpha, 1) \qquad (2)$$

This paper will heavily use the distribution of $|s_\alpha|^\alpha$:

$$|s_\alpha|^\alpha = \frac{|\sin(\alpha u)|^\alpha}{\cos u}\left[\frac{\cos(u - \alpha u)}{w}\right]^{(1-\alpha)} \qquad (3)$$

Intuitively, as $\alpha \to 0$, $1/|s_\alpha|^\alpha$ converges to $exp(1)$ in distribution as formally established by [7].

The use of α-stable distributions [9, 11] was studied in the context of estimating frequency moments of data streams [1, 17]. The use α-stable random projections for sparse signal recovery was established in (e.g.,) [16], by using full (i.e., infinite-bit) information of the measurements. In this paper, the development of binary (1-bit) and multi-bit coding schemes is also motivated by the work recent work on "one scan 1-bit compressed sensing" [12].

1.2 One Scan 1-Bit Compressed Sensing

In contrast to classical compressed sensing (CS) [8, 4] and 1-bit compressed sensing [3, 10, 18, 21], there is a recent line of work on sparse signal recovery based on heavy-tailed designs [16, 12]. The main algorithm of "one scan 1-bit compressed sensing" [12] is summarized in Algorithm 1. Given n measurements $y_j = \sum_{i=1}^N x_i s_{ij}$, $j = 1$ to n, where $s_{ij} \sim S(\alpha, 1)$ i.i.d. and x_i, $i = 1$ to N, is a sparse (and possibly dynamic/streaming) vector, the task is to recover x from only the signs of the measurements, i.e., $sign(y_j)$. Algorithm 1 provides a simple recipe for recovering x from $sign(y_j)$ by scanning the coordinates of the vector only once.

Algorithm 1 Stable measurement collection and the one scan 1-bit algorithm for sign recovery.

Input: K-sparse signal $\mathbf{x} \in \mathbb{R}^{1 \times N}$, design matrix $\mathbf{S} \in \mathbb{R}^{N \times M}$ with entries sampled from $S(\alpha, 1)$ with small α (e.g., $\alpha = 0.05$). We sample $u_{ij} \sim uniform(-\pi/2, \ \pi/2)$ and $w_{ij} \sim exp(1)$ and compute s_{ij} by (2).
Collect: Linear measurements: $y_j = \sum_{i=1}^N x_i s_{ij}$, $j = 1$ to M.
Compute: For each coordinate $i = 1$ to N, compute

$$Q_i^+ = \sum_{j=1}^M \log\left(1 + sgn(y_j)sgn(u_{ij})e^{-(K-1)w_{ij}}\right), \quad Q_i^- = \sum_{j=1}^M \log\left(1 - sgn(y_j)sgn(u_{ij})e^{-(K-1)w_{ij}}\right)$$

Output: For $i = 1$ to N, report the estimated sign: $\hat{sgn}(x_i) = \begin{cases} +1 & \text{if } Q_i^+ > 0 \\ -1 & \text{if } Q_i^- > 0 \\ 0 & \text{if } Q_i^+ < 0 \text{ and } Q_i^- < 0 \end{cases}$

This efficient recovery procedure, however, requires the knowledge of "K", which is the l_α norm $\sum_{i=1}^N |x_i|^\alpha$ as $\alpha \to 0+$. In practice, this K will typically have to estimated and the hope is that

we do not have to use too many additional measurements just for the task of estimating K. In this paper, we will elaborate that only 1 bit or a few bits per measurement can provide accurate estimates of K (as well as the general term $\sum_{i=1}^{N} |x_i|^\alpha$ for $0 < \alpha \leq 2$).

Because the samples y_j are heavy-tailed, using traditional estimators, the storage requirement for each sample can be substantial, which consequently would cause issues in data retrieval, transmission and decoding. It is thus very desirable if we just need 1 bit or a few bits for each $|y_j|$.

2 Estimation of Λ_α Using Full (Infinite-Bit) Information

Given n i.i.d. samples $y_j \sim S(\alpha, \Lambda_\alpha)$, $j = 1$ to n, we review various estimators of the scale parameter Λ_α using full information (i.e., infinite-bit). When $\alpha = 2$ (i.e., Gaussian), the *arithmetic mean* estimator is statistically optimal, i.e., the (asymptotic) variance reaches the reciprocal of the Fisher Information from classical statistics theory:

$$\hat{\Lambda}_{2,f} = \frac{1}{n}\sum_{j=1}^{n} |y_j|^2, \qquad Var\left(\hat{\Lambda}_{2,f}\right) = \frac{\Lambda_2^2}{n} 2 \tag{4}$$

When $\alpha = 1$, the MLE $\hat{\Lambda}_{1,f}$ is the solution to the equation

$$\sum_{j=1}^{n} \frac{\hat{\Lambda}_{1,f}^2}{\hat{\Lambda}_{1,f}^2 + y_j^2} = \frac{n}{2}, \quad Var\left(\hat{\Lambda}_{1,f}\right) = \frac{\Lambda_1^2}{n} 2 + O\left(\frac{1}{n^2}\right) \tag{5}$$

The *harmonic mean* estimator [11] is suitable for small α and becomes optimal as $\alpha \to 0+$:

$$\hat{\Lambda}_{\alpha,f,hm} = \frac{-\frac{2}{\pi}\Gamma(-\alpha)\sin\left(\frac{\pi}{2}\alpha\right)}{\sum_{j=1}^{n} |y_j|^{-\alpha}} \left(n - \left(\frac{-\pi\Gamma(-2\alpha)\sin(\pi\alpha)}{\left[\Gamma(-\alpha)\sin\left(\frac{\pi}{2}\alpha\right)\right]^2} - 1\right)\right) \tag{6}$$

$$Var\left(\hat{\Lambda}_{\alpha,f,hm}\right) = \frac{\Lambda_\alpha^2}{n}\left(\frac{-\pi\Gamma(-2\alpha)\sin(\pi\alpha)}{\left[\Gamma(-\alpha)\sin\left(\frac{\pi}{2}\alpha\right)\right]^2} - 1\right) + O\left(\frac{1}{n^2}\right) \tag{7}$$

where $\Gamma(.)$ is the gamma function. When $\alpha \to 0+$, the variance becomes $\frac{\Lambda_{0+}^2}{n} + O\left(\frac{1}{n^2}\right)$.

In summary, the optimal variances for $\alpha = 0+$, 1, and 2, are respectively

$$\frac{\Lambda_{0+}^2}{n} 1, \qquad \frac{\Lambda_1^2}{n} 2, \qquad \text{and} \qquad \frac{\Lambda_2^2}{n} 2 \tag{8}$$

Our goal is to develop 1-bit and multi-bit schemes to achieve variances which are close to be optimal.

3 1-Bit Coding and Estimation

Again, consider n i.i.d. samples $y_j \sim S(\alpha, \Lambda_\alpha)$, $j = 1$ to n. In this section, the task is to estimate Λ_α using just one bit information of each $|y_j|$, with a pre-determined threshold. To accomplish this, we consider a threshold C (which can be a function of α) and compare it with $|y_j|^\alpha$, $j = 1, 2, ..., n$. In other word, we store a "0" if $|y_j|^\alpha \leq C$ and a "1" if $|y_j|^\alpha > C$. Note that we can express $|y_j|^\alpha$ as

$$|y_j|^\alpha \sim \Lambda_\alpha |s_\alpha|^\alpha, \qquad s_\alpha \sim S(\alpha, 1).$$

Let f_α and F_α be the pdf and cdf of $|s_\alpha|^\alpha$, respectively. Then we can define p_1 and p_2 as follows

$$p_1 = \mathbf{Pr}\left(z_\alpha \leq C\right) = F_\alpha\left(C/\Lambda_\alpha\right),$$
$$p_2 = \mathbf{Pr}\left(z_\alpha > C\right) = 1 - p_1 = 1 - F_\alpha\left(C/\Lambda_\alpha\right)$$

which are needed for computing the likelihood. Denote

$$n_1 = \sum_{j=1}^{n} 1\{z_j \le C\}, \qquad n_2 = \sum_{j=1}^{n} 1\{z_j > C\}$$

The log-likelihood of the $n = n_1 + n_2$ observations is

$$l = n_1 \log p_1 + n_2 \log p_2 = n_1 \log F_\alpha\left(C/\Lambda_\alpha\right) + n_2 \log\left[1 - F_\alpha\left(C/\Lambda_\alpha\right)\right]$$

To seek the MLE (maximum likelihood estimator) of Λ_α, we need to compute the first derivative $l' = \frac{\partial l}{\partial \Lambda_\alpha}$:

$$l' = n_1 \frac{f_\alpha\left(C/\Lambda_\alpha\right)}{F_\alpha\left(C/\Lambda_\alpha\right)}\left(-\frac{C}{\Lambda_\alpha^2}\right) + n_2 \frac{-f_\alpha\left(C/\Lambda_\alpha\right)}{1 - F_\alpha\left(C/\Lambda_\alpha\right)}\left(-\frac{C}{\Lambda_\alpha^2}\right)$$

Setting $l' = 0$ yields the MLE solution denoted by $\hat{\Lambda}_\alpha$:

$$F_\alpha^{-1}\left(n_1/n\right) = C/\Lambda_\alpha \implies \hat{\Lambda}_\alpha = C/F_\alpha^{-1}\left(n_1/n\right)$$

To assess the estimation variance of $\hat{\Lambda}_\alpha$, we resort to classical theory of Fisher Information, which says

$$Var\left(\hat{\Lambda}_\alpha\right) = \frac{1}{-E\left(l''\right)} + O\left(\frac{1}{n^2}\right)$$

After some algebra, we obtain

$$E\left(l''\right) = -n\frac{C^2}{\Lambda_\alpha^4}\frac{f_\alpha^2}{F_\alpha(1 - F_\alpha)}$$

For convenience, we introduce $\eta = \frac{\Lambda_\alpha}{C}$, and we summarize the above results in Theorem 1, which also provides the exact expression of the $O\left(\frac{1}{n}\right)$ bias term using classical statistics results [2, 20].

Theorem 1 *Given n i.i.d. samples $y_j \sim S(\alpha, \Lambda_\alpha)$, $j = 1$ to n, a threshold C, and $n_1 = \sum_{j=1}^{n} 1\{z_j \le C\}$, the maximum likelihood estimator (MLE) of Λ_α is*

$$\hat{\Lambda}_\alpha = C/F_\alpha^{-1}\left(n_1/n\right) \tag{9}$$

Denote $\eta = \frac{\Lambda_\alpha}{C}$. The asymptotic bias of $\hat{\Lambda}_\alpha$ is

$$E\left(\hat{\Lambda}_\alpha\right) = \Lambda_\alpha + \frac{\Lambda_\alpha}{n}\frac{n_1}{n}\left(1 - \frac{n_1}{n}\right)\left(\frac{\eta^2}{f_\alpha^2(1/\eta)} + \frac{\eta f_\alpha'(1/\eta)}{2f_\alpha^3(1/\eta)}\right) + O\left(\frac{1}{n^2}\right) \tag{10}$$

and the asymptotic variance of $\hat{\Lambda}_\alpha$ is

$$Var\left(\hat{\Lambda}_\alpha\right) = \frac{\Lambda_\alpha^2}{n}V_\alpha\left(\eta\right) + O\left(\frac{1}{n^2}\right) \tag{11}$$

where

$$V_\alpha\left(\eta\right) = \eta^2 \frac{F_\alpha(1/\eta)(1 - F_\alpha(1/\eta))}{f_\alpha^2(1/\eta)}, \tag{12}$$

where f_α and F_α are the pdf and cdf of $|S(\alpha, 1)|^\alpha$, respectively, and $f_\alpha'(z) = \frac{\partial f_\alpha(z)}{\partial z}$.

Proof: See Appendix A. □

4

3.1 $\alpha \to 0+$

As $\alpha \to 0+$, we have $1/[s_\alpha]^\alpha \sim exp(1)$. Thus

$$F_{0+}(z) = e^{-1/z}, \quad f_{0+}(z) = \frac{1}{z^2}e^{-1/z}, \quad F_{0+}^{-1}(z) = \frac{1}{\log 1/z}$$

We can then derive the estimator and its variance as

$$\hat{\Lambda}_{0+} = \frac{C}{F_{0+}^{-1}(n_1/n)} = C\log n/n_1, \quad Var\left(\hat{\Lambda}_{0+}\right) = \frac{\Lambda_{0+}^2}{n}V_{0+}(\eta) + O\left(\frac{1}{n^2}\right)$$

where

$$V_{0+}(\eta) = \eta^2 \frac{F_\alpha(1/\eta)(1 - F_\alpha(1/\eta))}{f_\alpha^2(1/\eta)} = \frac{e^{-\eta} - e^{-2\eta}}{\eta^2 e^{-2\eta}} = \frac{e^\eta - 1}{\eta^2}$$

The minimum $V_{0+}(\eta)$ is 1.544, attained at $\eta = 1.594$. (In this paper, we keep 3 decimal places.)

3.2 $\alpha = 1$

By properties of Cauchy distribution, we know

$$F_1(z) = \frac{2}{\pi}\tan^{-1} z, \quad f_1(z) = \frac{2}{\pi}\frac{1}{1 + z^2}, \quad F_1^{-1}(z) = \tan\frac{\pi}{2}z$$

Thus, we can derive the estimator and variance

$$\hat{\Lambda}_1 = \frac{C}{\tan\frac{\pi}{2}\frac{n_1}{n}}, \quad Var\left(\hat{\Lambda}_1\right) = \frac{\Lambda_1^2}{n}V_1(\eta) + O\left(\frac{1}{n^2}\right)$$

The minimum of $V_1(\eta)$ is $\frac{\pi^2}{4}$, attained at $\eta = 1$. To see this, let $t = 1/\eta$. Then $V_1(\eta) = \frac{1}{t^2}\frac{F_1(t)(1-F_1(t))}{f_1^2(t)}$ and

$$\begin{aligned}
\frac{\partial \log V_1(\eta)}{\partial t} &= -\frac{2}{t} + \frac{f_1(t)}{F_1(t)} + \frac{-f_1(t)}{1 - F_1(t)} - 2\frac{f_1'(t)}{f_1(t)} \\
&= -\frac{2}{t} + \frac{4t}{1 + t^2} + \frac{\frac{1}{1+t^2}}{\tan^{-1} t} - \frac{\frac{2}{\pi}\frac{1}{1+t^2}}{1 - \frac{2}{\pi}\tan^{-1} t} \\
&= \frac{1}{1 + t^2}\left[t^2 - 1 + \frac{1}{\tan^{-1} t} - \frac{1}{\frac{\pi}{2} - \tan^{-1} t}\right]
\end{aligned}$$

Setting $\frac{\partial \log V_1(\eta)}{\partial t} = 0$, the solution is $t = 1$. Hence the optimum is attained at $\eta = 1$.

3.3 $\alpha = 2$

Since $S(2,1) \sim \sqrt{2} \times N(0,1)$, i.e., $|s_\alpha|^2 \sim 2\chi_1^2$, we have

$$F_2(z) = F_{\chi_1^2}(z/2), \quad f_2(z) = f_{\chi_1^2}(z/2)/2,$$

where $F_{\chi_1^2}$ and $f_{\chi_1^2}$ are the cdf and pdf of a chi-square distribution with 1 degree of freedom, respectively. The MLE is $\hat{\Lambda}_2 = \frac{C}{F_2^{-1}(n_1/n)}$ and the optimal variance of $\hat{\Lambda}_2$ is $\frac{\Lambda_2^2}{n}3.066$, attained at $\eta = \frac{\Lambda_2}{C} = 0.228$.

5

3.4   General $0 < \alpha \leq 2$

For general $0 < \alpha \leq 2$, the cdf F_α and pdf f_α can be computed numerically. Figure 1 plots $V_\alpha(\eta)$ for α from 0 to 2. The lowest point on each curve corresponds to the optimal (smallest) $V_\alpha(\eta)$. Figure 2 plots the optimal V_α values (left panel) and optimal η values (right panel).
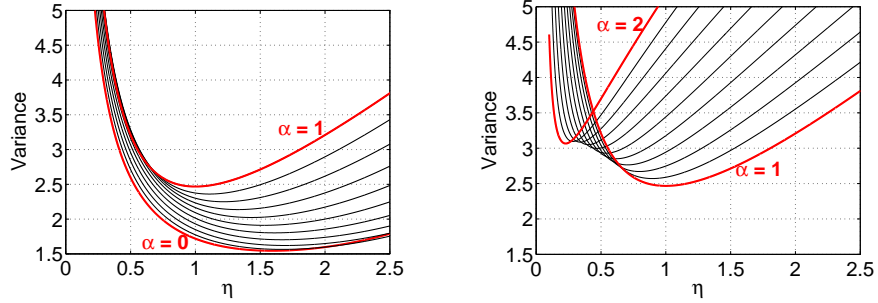

Figure 1: The variance factor $V_\alpha(\eta)$ in (12) for $\alpha \in [0,\ 2]$, spaced at 0.1. The lowest point on each curve corresponds to the optimal variance at that α value.
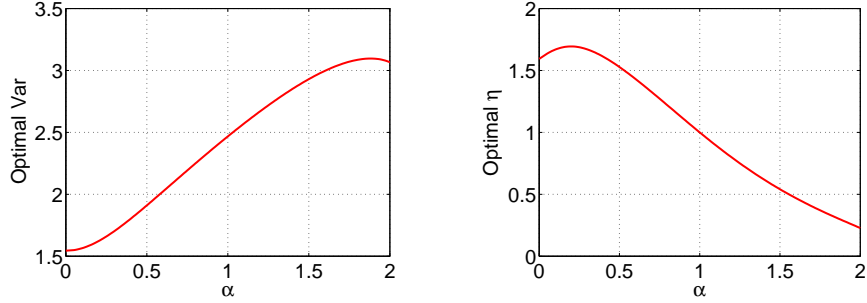

Figure 2: The optimal variance values $V_\alpha(\eta)$ (left panel) and the corresponding optimal η values (right panel). Each point on the curve corresponds to the lowest point of the curve for that α as in Figure 1.

Figure 1 suggests that the 1-bit scheme performs reasonably well. The optimal variance coefficient V_α is not much larger than the variance using full information. For example, when $\alpha = 1$, the optimal variance coefficient using full information is 2 (i.e., see (8)), while the optimal variance coefficient of the 1-bit scheme is just $\frac{\pi^2}{4} = 2.467$ which is only about 20% larger. Furthermore, we can see that, at least when $\alpha \leq 1$, $V_\alpha(\eta)$ is not very sensitive to η in a wide range of η values, which is practically important, because an optimal choice of η requires knowing Λ_α and is general not achievable. The best we can hope for is that the estimate is not sensitive to the choice of η.

3.5   Error Tail Bounds

Theorem 2

$$\mathbf{Pr}\left(\hat{\Lambda}_\alpha \geq (1+\epsilon)\Lambda_\alpha\right) \leq \exp\left(-n\frac{\epsilon^2}{G_{R,\alpha,C,\epsilon}}\right),\ \epsilon \geq 0$$

$$\mathbf{Pr}\left(\hat{\Lambda}_\alpha \leq (1-\epsilon)\Lambda_\alpha\right) \leq \exp\left(-n\frac{\epsilon^2}{G_{L,\alpha,C,\epsilon}}\right),\ 0 \leq \epsilon \leq 1$$

6

where $G_{R,\alpha,C,\epsilon}$ and $G_{L,\alpha,C,\epsilon}$ are computed as follows:

$$\frac{\epsilon^2}{G_{R,\alpha,C,\epsilon}} = - F_\alpha(1/(1+\epsilon)\eta) \log \left[\frac{F_\alpha(1/\eta)}{F_\alpha(1/(1+\epsilon)\eta)}\right] \tag{13}$$
$$- (1 - F_\alpha(1/(1+\epsilon)\eta)) \log \left[\frac{1 - F_\alpha(1/\eta)}{1 - F_\alpha(1/(1+\epsilon)\eta)}\right]$$

$$\frac{\epsilon^2}{G_{L,\alpha,C,\epsilon}} = - F_\alpha(1/(1-\epsilon)\eta) \log \left[\frac{F_\alpha(1/\eta)}{F_\alpha(1/(1-\epsilon)\eta)}\right] \tag{14}$$
$$- (1 - F_\alpha(1/(1-\epsilon)\eta)) \log \left[\frac{1 - F_\alpha(1/\eta)}{1 - F_\alpha(1/(1-\epsilon)\eta)}\right]$$

Proof: *See Appendix B. The proof is based on Chernoff's original tail bounds [6] for the binomial distribution.* □

To ensure the error $\mathbf{Pr}\left(\hat{\Lambda}_\alpha \geq (1+\epsilon)\Lambda_\alpha\right) + \mathbf{Pr}\left(\hat{\Lambda}_\alpha \leq (1-\epsilon)\Lambda_\alpha\right) \leq \delta, \;\; 0 \leq \delta \leq 1$, it suffices that

$$\exp\left(-n\frac{\epsilon^2}{G_{R,\alpha,C,\epsilon}}\right) + \exp\left(-n\frac{\epsilon^2}{G_{L,\alpha,C,\epsilon}}\right) \leq \delta \tag{15}$$

for which it suffices

$$n \geq \frac{G_{\alpha,C,\epsilon}}{\epsilon^2} \log 2/\delta, \qquad \text{where} \tag{16}$$
$$G_{\alpha,C,\epsilon} = \max\{G_{R,\alpha,C,\epsilon}, \; G_{L,\alpha,C,\epsilon}\}$$

Obviously, it will be even more precise to numerically compute n from (15) instead of using the convenient sample complexity bound (16). Figure 3 provides the tail bound constants for $\alpha = 0+$, i.e., $G_{R,0+,C,\epsilon}$ and $G_{L,0+,C,\epsilon}$ at selected η values ranging from 1 to 2.
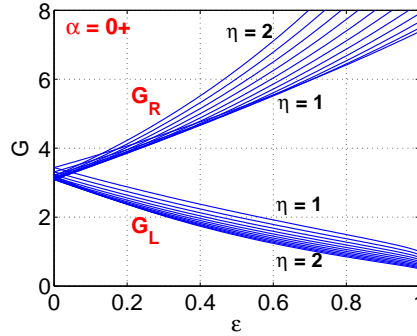

Figure 3: The tail bound constants $G_{R,0+,C,\epsilon}$ (13) (upper group) and $G_{L,0+,C,\epsilon}$ (14) (lower group), for $\eta = 1$ to 2 spaced at 0.1. Recall $\eta = \frac{\Lambda_\alpha}{C}$.

3.6 Bias-Correction

Bias-correction for MLE is important for small sample size n. In Theorem 1, Eq. (10) says

$$E\left(\hat{\Lambda}_\alpha\right) = \Lambda_\alpha + \frac{\Lambda_\alpha}{n}\frac{n_1}{n}\left(1 - \frac{n_1}{n}\right)\left(\frac{\eta^2}{f_\alpha^2(1/\eta)} + \frac{\eta f_\alpha'(1/\eta)}{2f_\alpha^3(1/\eta)}\right) + O\left(\frac{1}{n^2}\right)$$

7

which naturally provides a bias-correction for $\hat{\Lambda}_\alpha$, known as the "Bartlett correction" in statistics. To do so, we will need to use the estimate $\hat{\Lambda}_\alpha$ to compute the η. Since $\hat{\Lambda}_\alpha = C/F_\alpha^{-1}(n_1/n)$, we have $\hat{\Lambda}_\alpha/C = 1/F_\alpha^{-1}(n_1/n)$. The bias-corrected estimator, denoted by $\hat{\Lambda}_{\alpha,c}$ is

$$\hat{\Lambda}_{\alpha,c} = \frac{\hat{\Lambda}_\alpha}{1 + \frac{1}{n}\frac{n_1}{n}\left(1 - \frac{n_1}{n}\right)\left(\frac{\hat{\eta}^2}{f_\alpha^2(1/\hat{\eta})} + \frac{\hat{\eta}f_\alpha'(1/\hat{\eta})}{2f_\alpha^3(1/\hat{\eta})}\right)}, \quad \text{where} \quad \hat{\eta} = 1/F_\alpha^{-1}(n_1/n) \tag{17}$$

which, when $\alpha = 0+$, $\alpha = 1$, and $\alpha = 2$, becomes respectively

$$\hat{\Lambda}_{0+,c} = \frac{C\log n/n_1}{1 + \frac{1/n_1 - 1/n}{2\log n/n_1}} \tag{18}$$

$$\hat{\Lambda}_{1,c} = \frac{\frac{C}{\tan\frac{\pi}{2}\frac{n_1}{n}}}{1 + \frac{1}{n}\frac{\pi^2}{4}\frac{n_1}{n}\left(1 - \frac{n_1}{n}\right)\left(1 + \frac{1}{\tan^2\frac{\pi}{2}\frac{n_1}{n}}\right)} \tag{19}$$

$$\hat{\Lambda}_{2,c} = \frac{\frac{C}{2F_{\chi_1^2}^{-1}(n_1/n)}}{1 + \frac{\pi}{n}\frac{n_1}{n}\left(1 - \frac{n_1}{n}\right)\left(\frac{3}{F_{\chi_1^2}^{-1}(n_1/n)} - 1\right)e^{F_{\chi_1^2}^{-1}(n_1/n)}} \tag{20}$$

See the detailed derivations in Appendix A, together with the proof of Theorem 1.

4 Experiments on 1-Bit Coding and Estimation

We conduct extensive simulations to (i) verify the 1-bit variance formulas of the MLE, and (ii) apply the 1-bit estimator in Algorithm 1 for one scan 1-bit compressed sensing [12].

4.1 Bias and Variance of the Proposed Estimators

Figure 4 provides the simulations for verifying the 1-bit estimator $\hat{\Lambda}_{0+}$ and its bias-corrected version $\hat{\Lambda}_{0+,c}$ using small α (i.e., 0.05). Basically, for each sample size n, we generate 10^6 samples from $S(\alpha, 1)$, which are quantized according a pre-selected threshold C. Then we apply both $\hat{\Lambda}_{0+}$ and $\hat{\Lambda}_{0+,c}$ and report the empirical mean square error (MSE = variance + bias^2) from 10^6 repetitions. For thorough evaluations, we conduct simulations for a wide range of $n \in [5, 1000]$.

The results are presented in log-log scale, which exaggerates the portion for small n and the y-axis for large n. The plots confirm that when n is not too small (e.g., $n > 100$), the bias of MLE estimate varnishes and the asymptotic variance formula (12) matches the mean square error. For small n (e.g., $n < 100$), the bias correction becomes important.

Note that when n is large (i.e., when errors are very small), the plots show some discrepancies. This is due to the fact that we have to use small α for the simulations but the estimators $\hat{\Lambda}_{0+}$ and $\hat{\Lambda}_{0+,c}$ are based on $\alpha = 0+$. The differences are very small and only become visible when the estimation errors are so small (due to the exaggeration of the log-scale). To remove this effect, we also conduct similar simulations for $\alpha = 1$ and present the results in Figure 5, which does not show the discrepancies at large n. We can see that the bias-correction step is also important for $\alpha = 1$.

We should mention that, for numerical issue, we added a small real number (10^{-6}) to n_1. We did not further investigate various smoothing techniques as it appears that this Bartlett-correction procedure already serves the purpose well.
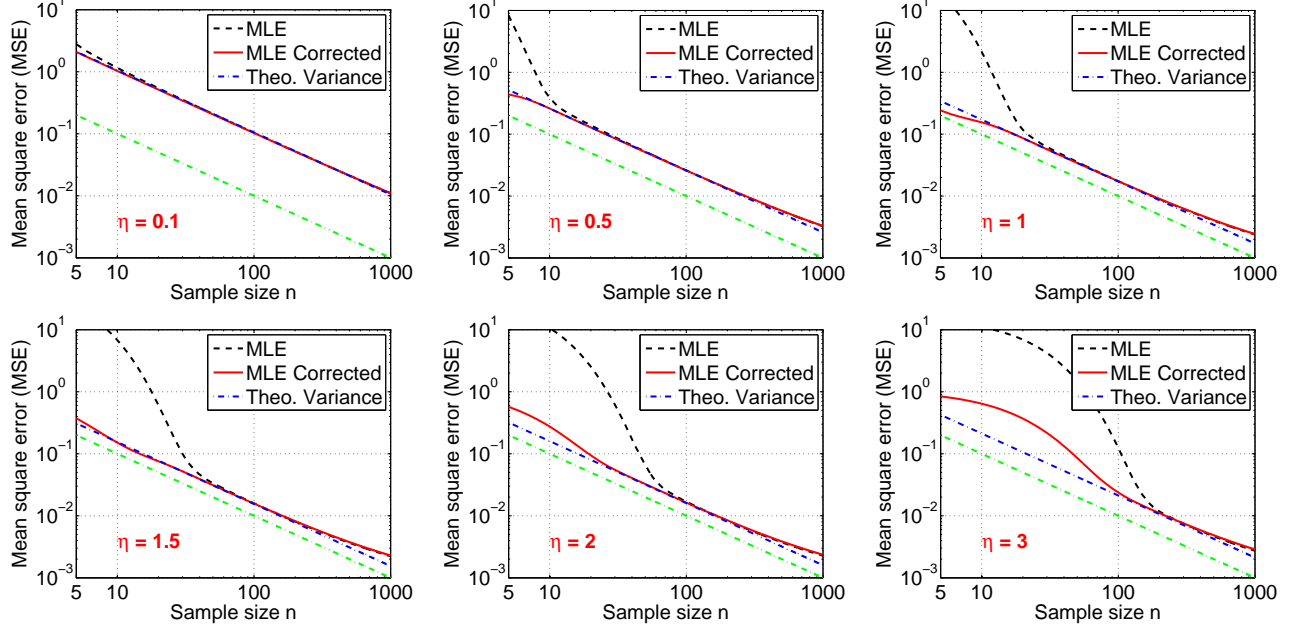
8

Figure 4: Empirical Mean square errors of $\hat{\Lambda}_{0+}$ (dashed curves) and $\hat{\Lambda}_{0+,c}$ (solid curves) from 10^6 simulations of $S(\alpha, 1)$ for $\alpha = 0.05$, at each sample size n. Each panel present results for a different $\eta = \frac{\Lambda_\alpha}{C}$. For both estimators, the empirical MSEs converge to the theoretical asymptotic variances (12) (dashed dot curves and blue if color is available) when n is large enough. In each panel, the lowest curve (dashed dot and green if color is available) represents the theoretical variance using full (infinite-bit) information, i.e., $1/n$ in this case. For small n, the bias-correction step important. Note that the small (and exaggerated) discrepancies at large n are due to use of $\alpha = 0.05$ to generate samples and the estimators based on $\alpha = 0+$. Also recall that $\eta = 1.594$ is the optimal η.
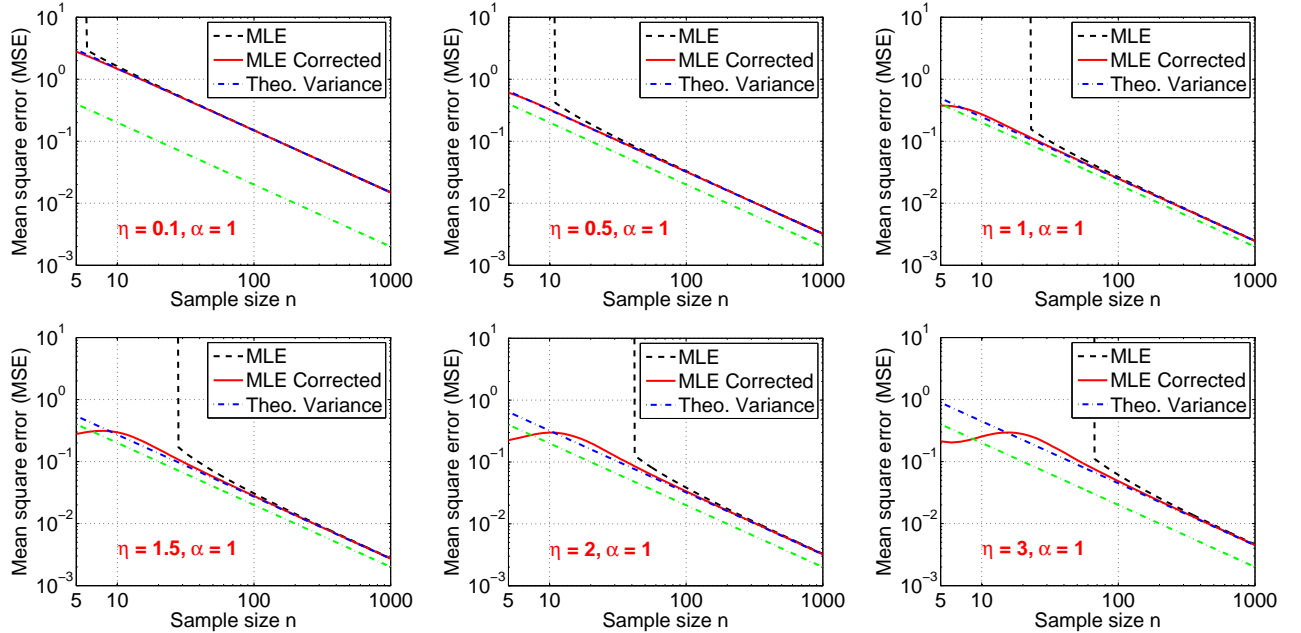

Figure 5: Mean square errors of $\hat{\Lambda}_1$ (dashed curves) and $\hat{\Lambda}_{1,c}$ (solid curves) for $\alpha = 1$. Note that the lowest curve (dashed dot and green if color is available) in each panel represents the optimal variance using full (i.e., infinite-bit) information, which is $2/n$ for $\alpha = 1$.

9

4.2 One Scan 1-Bit Compressed Sensing

Next, we integrate $\hat{\Lambda}_{0+,c}$ into the sparse recovery procedure in Algorithm 1, by replacing K with $\hat{\Lambda}_{0+,c}$ for computing Q_i^+ and Q_i^- [12]. We report the sign recovery errors $\sum_i |\hat{sgn}(x_i) - sgn(x_i)|/K$ from 10^4 simulations. In this study, we let $N = 1000$, $K = 20$, and sample the nonzero coordinates from $N(0, 5^2)$. For estimating K, we use $n \in \{20, 50, 100\}$ samples with $\eta \in \{0.2, 0.5, 1.5, 2, 3\}$. Recall $\eta = 1.5$ is close to be optimal (1.594) for $\hat{\Lambda}_{0+}$.

Figure 6 reports the sign recovery errors at 75% quantile (upper panels) and 95% quantile (bottom panels). The number of measurements for sparse recovery is chosen according to $M = \zeta K \log(N/0.01)$, although we only use $n \in \{20, 50, 100\}$ samples to estimate K. For comparison, Figure 6 also reports the results for estimating K using n full (i.e., infinite-bit) samples.

When $n = 100$, except for $\eta = 0.2$ (which is too small), the performance of $\hat{\Lambda}_{0+,c}$ is fairly stable with no essential difference from the estimator using full information. The performance of $\hat{\Lambda}_{0+,c}$ deteriorates with decreasing n. But even for $n = 20$, $\hat{\Lambda}_{0+,c}$ at $\eta = 1.5$ still performs well.
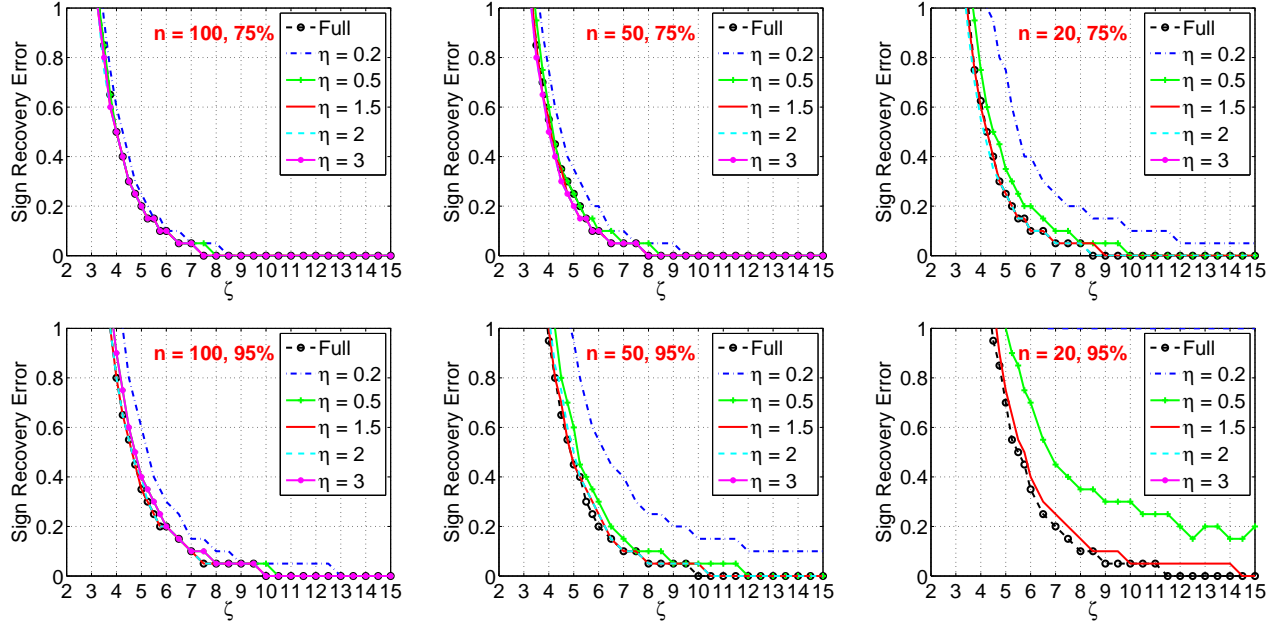

Figure 6: Sign recovery error: $\sum_i |\hat{sgn}(x_i) - sgn(x_i)|/K$, using Algorithm 1 and estimated K in computing Q_i^+ and Q_i^- in Algorithm 1. In this study, $N = 1000$, $K = 20$, and the nonzero entries are generated from $N(0, 5^2)$. The number of measurements for recovery is $M = \zeta K \log(N/0.01)$ and we use n samples to estimate K for $n \in \{20, 50, 100\}$. We report 75% (upper panels) and 95% (bottom panels) quantiles of the sign recovery errors, from 10^4 repetitions. We estimate K using the full information (i.e., the estimator (6)) as well as 1-bit estimator $\hat{\Lambda}_{0+,c}$ with selected values of $\eta \in \{0.2, 0.5, 1.5, 2, 3\}$. When $n = 100$, except for $\eta = 0.2$ (which is too small), the performance of $\hat{\Lambda}_{0+,c}$ is fairly stable with no essential difference from the estimator using full information. The performance of $\hat{\Lambda}_{0+,c}$ deteriorates with decreasing n. But even when $n = 20$, the performance of $\hat{\Lambda}_{0+,c}$ at $\eta = 1.5$ (which is close to be optimal) is still very good. Note that, when a curve does not show in the panel (e.g., $n = 50$, $\eta = 3$, and 95%), it basically means the error is too large to fit in.

5 2-Bit Coding and Estimation

As shown by theoretical analysis and simulations, the performance of 1-bit coding and estimation is fairly good and stable for a wide range of threshold values. Nevertheless, it is desirable to further stabilize the estimates (and lower the variance) by using more bits.

With the 2-bit scheme, we need to introduce 3 threshold values: $C_1 \leq C_2 \leq C_3$. We define

$$
\begin{aligned}
p_1 &= \mathbf{Pr}\left(z_j \leq C_1\right) = F_\alpha\left(C_1/\Lambda_\alpha\right) \\
p_2 &= \mathbf{Pr}\left(C_1 < z_j \leq C_2\right) = F_\alpha\left(C_2/\Lambda_\alpha\right) - F_\alpha\left(C_1/\Lambda_\alpha\right) \\
p_3 &= \mathbf{Pr}\left(C_2 < z_j \leq C_3\right) = F_\alpha\left(C_3/\Lambda_\alpha\right) - F_\alpha\left(C_2/\Lambda_\alpha\right) \\
p_4 &= \mathbf{Pr}\left(z_j > C_3\right) = 1 - F_\alpha\left(C_3/\Lambda_\alpha\right)
\end{aligned}
$$

and

$$
n_1 = \sum_{j=1}^{n} 1\{z_j \leq C_1\}, \quad n_2 = \sum_{j=1}^{n} 1\{C_1 < z_j \leq C_2\}
$$

$$
n_3 = \sum_{j=1}^{n} 1\{C_2 < z_j \leq C_3\}, \quad n_4 = \sum_{j=1}^{n} 1\{z_j > C_3\}
$$

The log-likelihood of these $n = n_1 + n_2 + n_3 + n_4$ observations can be expressed as

$$
\begin{aligned}
l =&n_1 \log p_1 + n_2 \log p_2 + n_3 \log p_3 + n_4 \log p_4 \\
=&n_1 \log F_\alpha\left(C_1/\Lambda_\alpha\right) + n_2 \log\left[F_\alpha\left(C_2/\Lambda_\alpha\right) - F_\alpha\left(C_1/\Lambda_\alpha\right)\right] + \\
&n_3 \log\left[F_\alpha\left(C_3/\Lambda_\alpha\right) - F_\alpha\left(C_2/\Lambda_\alpha\right)\right] + n_4 \log\left[1 - F_\alpha\left(C_3/\Lambda_\alpha\right)\right],
\end{aligned}
$$

from which we can derive the MLE and variance as presented in Theorem 3.

Theorem 3 *Given n i.i.d. samples $y_j \sim S(\alpha, 1)$, $j = 1$ to n, three thresholds $0 < C_1 \leq C_2 \leq C_3$, $n_1 = \sum_{j=1}^{n} 1\{z_j \leq C_1\}$, $n_2 = \sum_{j=1}^{n} 1\{C_1 < z_j \leq C_2\}$, $n_3 = \sum_{j=1}^{n} 1\{C_2 < z_j \leq C_3\}$, $n_4 = \sum_{j=1}^{n} 1\{z_j > C_3\}$, and*

$$
\eta_1 = \frac{\Lambda_\alpha}{C_1}, \qquad \eta_2 = \frac{\Lambda_\alpha}{C_2}, \qquad \eta_3 = \frac{\Lambda_\alpha}{C_3}
$$

the MLE, denoted by $\hat{\Lambda}_\alpha$, is the solution to the following equation:

$$
\begin{aligned}
0 =&n_1 \frac{C_1 f_\alpha\left(1/\eta_1\right)}{F_\alpha\left(1/\eta_1\right)} + n_2 \frac{C_2 f_\alpha\left(1/\eta_2\right) - C_1 f_\alpha\left(1/\eta_1\right)}{F_\alpha\left(1/\eta_2\right) - F_\alpha\left(1/\eta_1\right)} \\
&+ n_3 \frac{C_3 f_\alpha\left(1/\eta_3\right) - C_2 f_\alpha\left(1/\eta_2\right)}{F_\alpha\left(1/\eta_3\right) - F_\alpha\left(1/\eta_2\right)} + n_4 \frac{-C_3 f_\alpha\left(1/\eta_3\right)}{1 - F_\alpha\left(1/\eta_3\right)}
\end{aligned}
$$

The asymptotic variance of the MLE is

$$
Var\left(\hat{\Lambda}_\alpha\right) = \frac{\Lambda_\alpha^2}{n} V_\alpha(\eta_1, \eta_2, \eta_3) + O\left(\frac{1}{n^2}\right)
$$

where the variance factor can be expressed as

$$
\begin{aligned}
\frac{1}{V_\alpha(\eta_1, \eta_2, \eta_3)} =&\frac{1}{\eta_1^2} \frac{f_\alpha^2\left(1/\eta_1\right)}{F_\alpha\left(1/\eta_1\right)} + \frac{1}{\eta_3^2} \frac{f_\alpha^2\left(1/\eta_3\right)}{1 - F_\alpha\left(1/\eta_3\right)} + \frac{\left[f_\alpha\left(1/\eta_2\right)/\eta_2 - f_\alpha\left(1/\eta_1\right)/\eta_1\right]^2}{F_\alpha\left(1/\eta_2\right) - F_\alpha\left(1/\eta_1\right)} \\
&+ \frac{\left[f_\alpha\left(1/\eta_3\right)/\eta_3 - f_\alpha\left(1/\eta_2\right)/\eta_2\right]^2}{F_\alpha\left(1/\eta_3\right) - F_\alpha\left(1/\eta_2\right)}
\end{aligned}
$$

The asymptotic bias is

$$E\left(\hat{\Lambda}_\alpha\right) = \Lambda_\alpha\left(1 + \frac{1}{nB} - \frac{D}{2nB^2}\right) + O\left(\frac{1}{n^2}\right)$$

where

$$B = \frac{\left(-\frac{C_1}{\Lambda_\alpha}\right)^2 f_1^2}{F_1} + \frac{\left[\left(-\frac{C_2}{\Lambda_\alpha}\right)f_2 - \left(-\frac{C_1}{\Lambda_\alpha}\right)f_1\right]^2}{F_2 - F_1} + \frac{\left[\left(-\frac{C_3}{\Lambda_\alpha}\right)f_3 - \left(-\frac{C_2}{\Lambda_\alpha}\right)f_2\right]^2}{F_3 - F_2} + \frac{\left(-\frac{C_3}{\Lambda_\alpha}\right)^2 f_3^2}{1 - F_3}$$

and

$$D = \frac{\left(-\frac{C_1}{\Lambda_\alpha}\right)^3 f_1 f_1'}{F_1} + \frac{\left[\left(-\frac{C_2}{\Lambda_\alpha}\right)f_2 - \left(-\frac{C_1}{\Lambda_\alpha}\right)f_1\right]\left[\left(-\frac{C_2}{\Lambda_\alpha}\right)^2 f_2' - \left(-\frac{C_1}{\Lambda_\alpha}\right)^2 f_1'\right]}{F_2 - F_1}$$
$$+ \frac{\left[\left(-\frac{C_3}{\Lambda_\alpha}\right)f_3 - \left(-\frac{C_2}{\Lambda_\alpha}\right)f_2\right]\left[\left(-\frac{C_3}{\Lambda_\alpha}\right)^2 f_3' - \left(-\frac{C_2}{\Lambda_\alpha}\right)^2 f_2'\right]}{F_3 - F_2} + \frac{\left(-\frac{C_3}{\Lambda_\alpha}\right)^3 f_3 f_3'}{1 - F_3}$$

Proof: *See Appendix C.*   □

The asymptotic bias formula in Theorem 3 leads to a bias-corrected estimator

$$\hat{\Lambda}_{\alpha,c} = \frac{\hat{\Lambda}_\alpha}{1 + \frac{1}{nB} - \frac{D}{2nB^2}} \tag{21}$$

Note that, with a slight abuse of notation, we still use $\hat{\Lambda}_\alpha$ to denote the MLE of the 2-bit scheme and we rely on the number of parameters (e.g., η_1, η_2, η_3) to differentiate V_α for different schemes.

5.1 $\alpha \to 0+$

In this case, we can slightly simplify the expression:

$$V_\alpha(\eta_1, \eta_2, \eta_3) = \frac{1}{\frac{(\eta_1-\eta_2)^2}{e^{\eta_1}-e^{\eta_2}} + \frac{(\eta_2-\eta_3)^2}{e^{\eta_2}-e^{\eta_3}} + \frac{\eta_3^2}{e^{\eta_3}-1}}$$

Numerically, the minimum of $V_{0+}(\eta_1, \eta_2, \eta_3)$ is 1.122, attained at $\eta_1 = 3.365, \eta_2 = 1.771, \eta_3 = 0.754$. The value 1.122 is substantially smaller than 1.544 which is the minimum variance coefficient of the 1-bit scheme. Figure 7 illustrates that, with the 2-bit scheme, the variance is less sensitive to the choice of the thresholds, compared to the 1-bit scheme.

In practice, there are at least two simple strategies for selecting the parameters $\eta_1 \geq \eta_2 \geq \eta_3$:

- *Strategy 1*: First select a "small" η_3, then let $\eta_2 = t\eta_3$ and $\eta_1 = t\eta_2$, for some $t > 1$.

- *Strategy 2*: First select a "small" η_3 and a "large" η_1, then select a "reasonable" η_2 in between.

See the plots for examples of the two strategies in Figure 7. We re-iterate that for the task of estimating Λ_α using only a few bits, we must choose parameters (thresholds) beforehand. While in general the optimal results are not attainable, as long as the chosen parameters fall in a "reasonable" range (which is fairly wide), the estimation variance will not be far away from the optimal value.
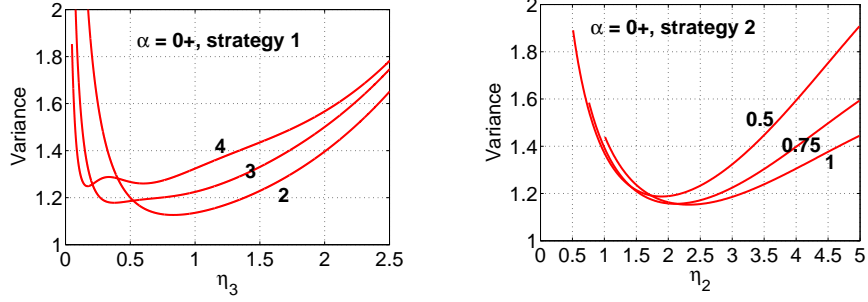
Figure 7: **Left** (strategy 1): $V_{0+}(\eta_1, \eta_2, \eta_3)$ for $\eta_2 = t\eta_3$, $\eta_1 = t\eta_2$, at $t = 2, 3, 4$, with varying η_3. **Right** (strategy 2): V_{0+} for fixed $\eta_1 = 5$, $\eta_3 \in \{0.5,\ 0.75,\ 1\}$, and η_2 varying between η_3 and η_1.

5.2 $\quad \alpha = 1$

Numerically, the minimum of $V_1(\eta_1, \eta_2, \eta_3)$ is 2.087, attained at $\eta_1 = 1.927$, $\eta_2 = 1.000$, $\eta_3 = 0.519$. Note that the value 2.087 is very close to the optimal variance coefficient 2 using full information. Figure 8 plots the examples of $V_1(\eta_1, \eta_2, \eta_3)$ for both "strategy 1" and "strategy 2".
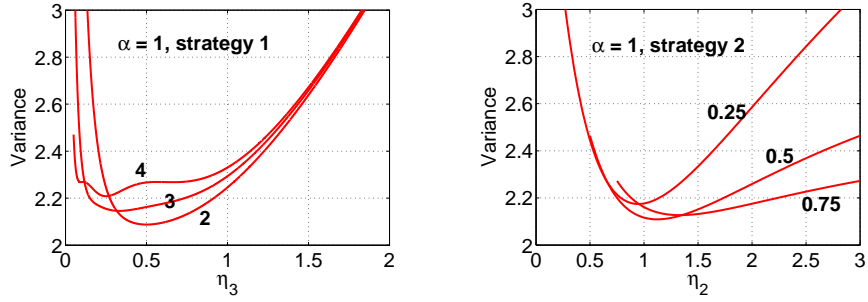

Figure 8: **Left** (strategy 1): $V_1(\eta_1, \eta_2, \eta_3)$ for $\eta_2 = t\eta_3$, $\eta_1 = t\eta_2$, at $t = 2, 3, 4$, with varying η_3. **Right** (strategy 2): V_1 for fixed $\eta_1 = 3$, $\eta_3 \in \{0.25,\ 0.5,\ 0.75\}$, and η_2 varying between η_3 and η_1.

5.3 $\quad \alpha = 2$

Numerically, the minimum of $V_2(\eta_1, \eta_2, \eta_3)$ is 2.236, attained at $\eta_1 = 0.546$, $\eta_2 = 0.195$, $\eta_3 = 0.093$. Figure 9 presents examples of $V_2(\eta_1, \eta_2, \eta_3)$ for both strategies for choosing η_1, η_2, and η_3.
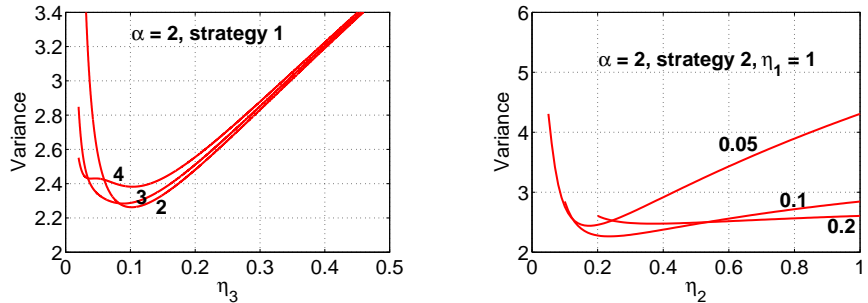

Figure 9: **Left** (strategy 1): $V_1(\eta_1, \eta_2, \eta_3)$ for $\eta_2 = t\eta_3$, $\eta_1 = t\eta_2$, at $t = 2, 3, 4$, with varying η_3. **Right** (strategy 2): V_1 for fixed $\eta_1 = 1$, $\eta_3 \in \{0.05,\ 0.1,\ 0.2\}$, and η_2 varying between η_3 and η_1.

5.4 Simulations

Figure 10 presents the simulation results for verifying the 2-bit estimator $\hat{\Lambda}_{0+}$ and its bias-corrected version $\hat{\Lambda}_{0+,c}$. For simplicity, we choose $\eta_3 \in \{0.05,\ 0.1,\ 0.25,\ 0.75,\ 1.5,\ 2\}$ and we fix $\eta_2 = 3\eta_3$, $\eta_1 = 3\eta_2$. Although these choices are not optimal, we can see from Figure 10 that the estimators still perform well for such a wide range of η_3 values. Compared to 1-bit estimators, the 2-bit estimators are noticeably more accurate and less sensitive to parameters. Again, the bias-correction step is useful when the sample size n is not large.

Similar to Figure 4, we can observe some discrepancies at large n (as magnified by the log-scale of the y-axis). Again, this is because we simulate the data using $\alpha = 0.05$ and we use estimators based on $\alpha = 0+$. To remove this effect, we also provide simulations for $\alpha = 1$ in Figure 11.
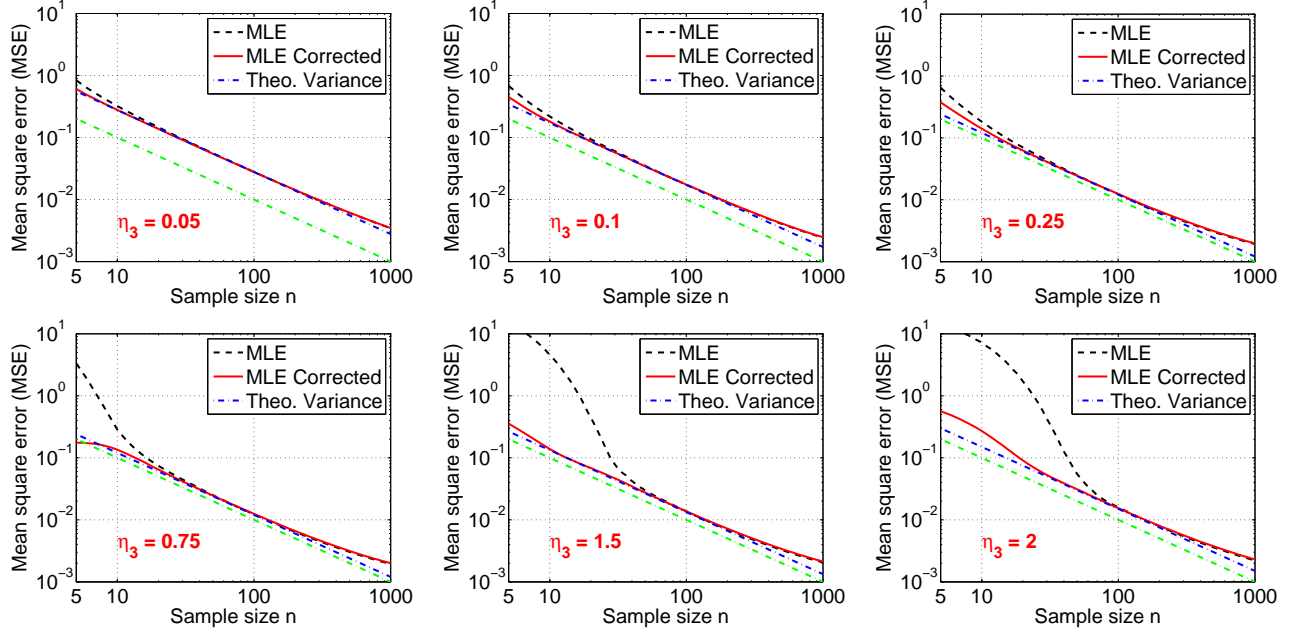

Figure 10: Empirical Mean square errors of the 2-bit estimators: $\hat{\Lambda}_{0+}$ (dashed curves) and $\hat{\Lambda}_{0+,c}$ (solid curves), for 10^6 simulations at each sample size n. We use $\alpha = 0.05$ to generate stable samples $S(\alpha, 1)$ and we consider 6 different $\eta_3 = \frac{\Lambda_\alpha}{C_3}$ values presented in 6 panels. We always let $\eta_2 = 3\eta_3$ and $\eta_1 = 3\eta_2$. For both estimators, the empirical MSEs converge to the theoretical asymptotic variances (12) (dashed dot curves and blue if color is available) when n is not small. In each panel, the lowest curve (dashed dot and green if color is available) represents the theoretical variances using full (infinite-bit) information, i.e., $1/n$ in this case. When n is small, the bias-correction step important. Note that the small (and exaggerated) discrepancies at large n are due to the fact that we use $\alpha = 0.05$ to simulate the data and use estimators based on $\alpha = 0+$.
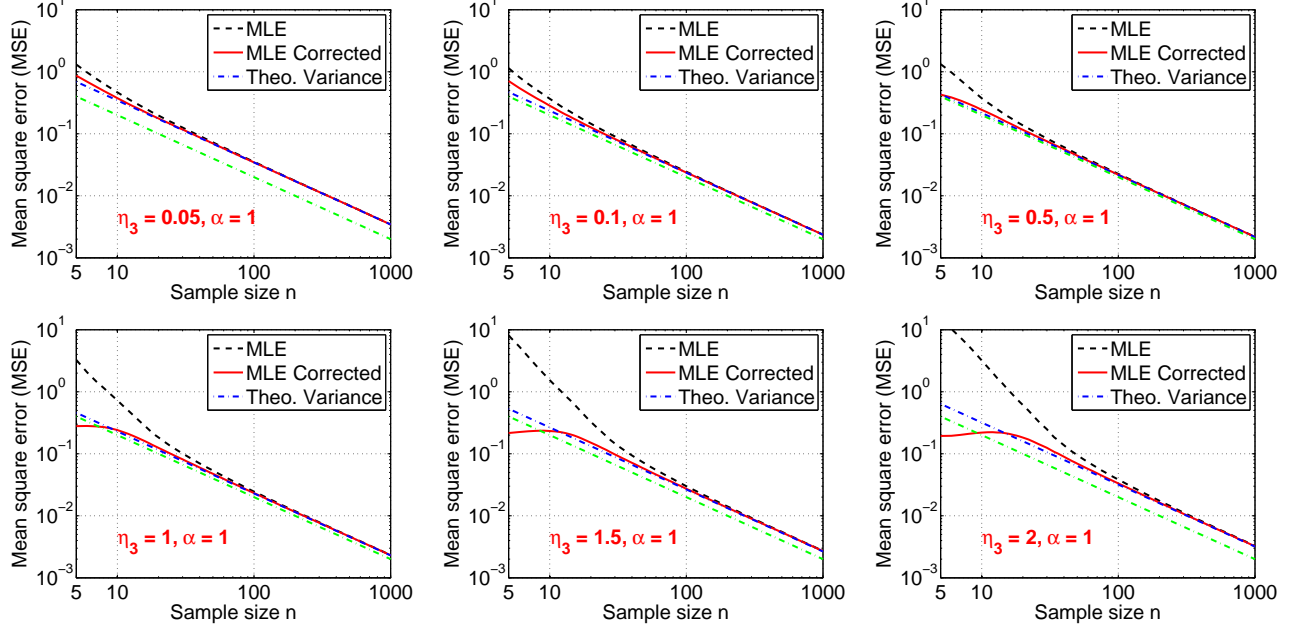
14

Figure 11: Mean square errors of the 2-bit estimator $\hat{\Lambda}_1$ (dashed curves) and its bias-corrected version $\hat{\Lambda}_{1,c}$ (solid curves), for $\alpha = 1$, by using 6 different η_3 values (one for each panel) and fixing $\eta_2 = 3\eta_3$, $\eta_1 = 3\eta_2$. The lowest curve (dashed dot and green if color is available) in each panel represents the optimal variance using full information, which is $2/n$ for $\alpha = 1$.

5.5 Efficient Computational Procedure for the MLE Solutions

With the 1-bit scheme, the cost for computing the MLE is negligible because of the closed-form solution. With the 2-bit scheme, however, the computational cost might be a concern if we try to find the MLE solution numerically every time (at run time). A computationally efficient solution is to tabulate the results. To see this, we can re-write the log-likelihood function

$$l = \frac{n_1}{n} \log F_\alpha \left(1/\eta_1 \right) + \frac{n_2}{n} \log \left[F_\alpha \left(1/\eta_2 \right) - F_\alpha \left(1/\eta_1 \right) \right]$$
$$+ \frac{n_3}{n} \log \left[F_\alpha \left(1/\eta_3 \right) - F_\alpha \left(\eta_2 \right) \right] + \frac{n - (n_1 + n_2 + n_3)}{n} \log \left[1 - F_\alpha \left(\eta_3 \right) \right]$$

This means, we only need to tabulate the results for the combination of $n_1/n, n_2/n, n_3/n$ (which all vary between 0 and 1). Suppose we tabulate T values for each n_i/n (i.e., at an accuracy of $1/T$), then the table size is only T^3, which is merely 10^6 if we let $T = 100$.

Here we conduct a simulation study for $\alpha = 1$ and $T \in \{20, 50, 100, 200\}$, as presented in Figure 12. We let $\eta_3 = 0.5$, $\eta_2 = 3\eta_3$, $\eta_1 = 3\eta_2$. We can see that the results are already good when $T = 100$ (or even just $T = 50$). This confirms the effectiveness of the tabulation scheme.

Therefore, tabulation provides an efficient solution to the computational problem for finding the MLE. Here, we have presented only a simple tabulation scheme based on uniform grids. It is possible to improve the scheme by using, for example, adaptive grids.
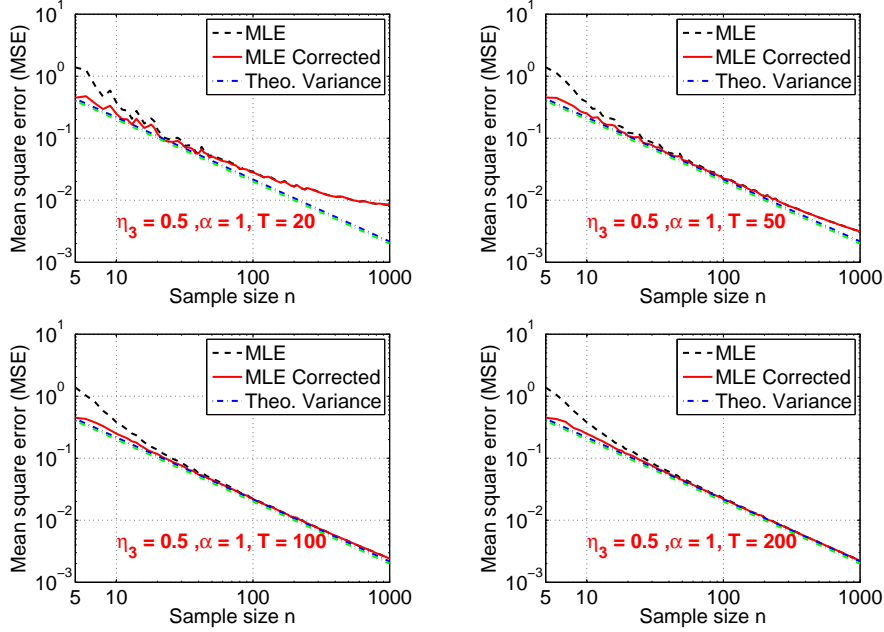
15

Figure 12: Mean square errors of the 2-bit tabulation-based estimator $\hat{\Lambda}_1$ (dashed curves) and its bias-corrected version $\hat{\Lambda}_{1,c}$ (solid curves), for $\alpha = 1$ and $T \in \{20, 50, 100, 200\}$ tabulation levels. by fixing $\eta_3 = 0.5$, $\eta_2 = 3\eta_3$, $\eta_1 = 3\eta_2$. The lowest curve (dashed dot and green if color is available) in each panel represents the optimal variance using full information, which is $2/n$ for $\alpha = 1$.

6 Multi-Bit (Multi-Partition) Coding and Estimation

Clearly, we can extend this methodology to more than 2 bits. With more bits, it is more flexible to consider schemes based on $(m + 1)$ partitions. For example $m = 1$ for the 1-bit scheme, $m = 3$ for the 2-bit scheme, and $m = 7$ for the 3-bit scheme. We feel $m \leq 5$ is practical. The asymptotic variance of the MLE $\hat{\Lambda}_\alpha$ can be expressed as

$$Var\left(\hat{\Lambda}_\alpha\right) = \frac{\Lambda_\alpha^2}{n} V_\alpha(\eta_1, ..., \eta_m) + O\left(\frac{1}{n^2}\right), \qquad \text{where}$$

$$\frac{1}{V_\alpha(\eta_1, ..., \eta_m)} = \frac{1}{\eta_1^2} \frac{f_\alpha^2(1/\eta_1)}{F_\alpha(1/\eta_1)} + \frac{1}{\eta_m^2} \frac{f_\alpha^2(1/\eta_m)}{1 - F_\alpha(1/\eta_m)} + \sum_{s=1}^{m-1} \frac{\left[f_\alpha(1/\eta_{s+1})/\eta_{s+1} - f_\alpha(1/\eta_s)/\eta_s\right]^2}{F_\alpha(1/\eta_{s+1}) - F_\alpha(1/\eta_s)}$$

Here, we provide some numerical results for $m = 5$, to demonstrate that using more partitions does further reduce the estimation variances and further stabilize the estimates in that the estimation accuracy is not as sensitive to parameters.

6.1 $\alpha = 0+$ and $m = 5$

Numerically, the minimum of $V_{0+}(\eta_1, \eta_2, \eta_3, \eta_4, \eta_5)$ is 1.055, attained at $\eta_1 = 4.464$, $\eta_2 = 2.871$, $\eta_3 = 1.853$, $\eta_4 = 1.099$, $\eta_5 = 0.499$. Figure 13 (right panel) plots $V_{0+}(\eta_1, \eta_2, \eta_3, \eta_4, \eta_5)$ for varying η_5 and $\eta_i = t\eta_{i+1}$, $i = 4, 3, 2, 1$. For comparison, we also plot (in the left panel) $V_{0+}(\eta_1, \eta_2, \eta_3)$ for varying η_3, and $\eta_2 = t\eta_3$, $\eta_1 = t\eta_2$. We can see that with more partitions, the performance becomes significantly more robust.
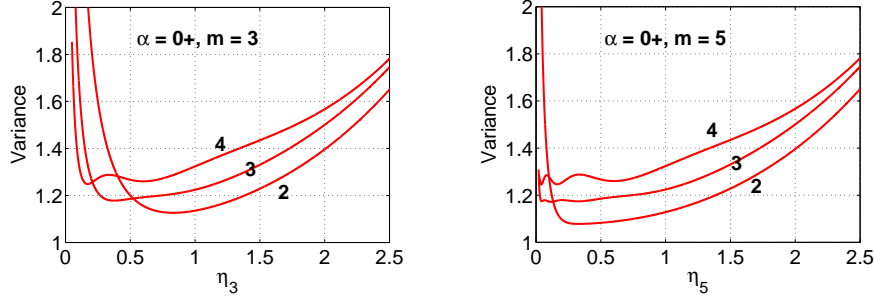
Figure 13: **Left** (4-partition): $V_{0+}(\eta_1, \eta_2, \eta_3)$ for varying η_3 and $\eta_2 = t\eta_3$, $\eta_1 = t\eta_2$, at $t = 2, 3, 4$. **Right** (6-partition): $V_{0+}(\eta_1, \eta_2, \eta_3, \eta_4, \eta_5)$ for varying η_5 and $\eta_i = t\eta_{i+1}$, at $t = 2, 3, 4$.

6.2 $\alpha = 1$ and $m = 5$

Numerically, the minimum of $V_1(\eta_1, \eta_2, \eta_3, \eta_4, \eta_5)$ is 2.036, attained at $\eta_1 = 2.602$, $\eta_2 = 1.498$, $\eta_3 = 1.001$, $\eta_4 = 0.668$, $\eta_5 = 0.385$. Figure 14 (right panel) plots $V_1(\eta_1, \eta_2, \eta_3, \eta_4, \eta_5)$ for varying η_5 and $\eta_i = t\eta_{i+1}$, $i = 4, 3, 2, 1$. Again, for comparison, we also plot (in the left panel) $V_1(\eta_1, \eta_2, \eta_3)$ for varying η_3, and $\eta_2 = t\eta_3$, $\eta_1 = t\eta_2$. Clearly, using more partitions stabilizes the variances even when the parameters are chosen less optimally.
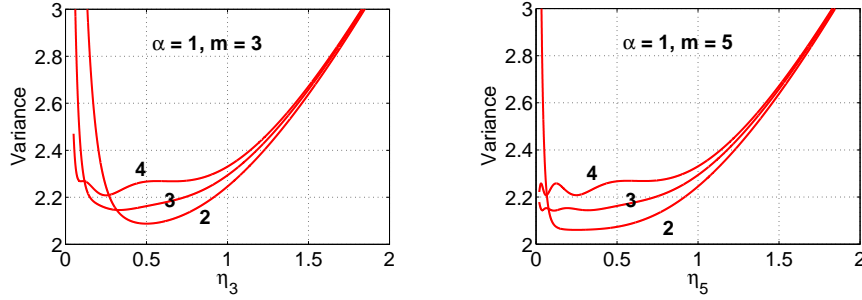

Figure 14: **Left** (4-partition): $V_1(\eta_1, \eta_2, \eta_3)$ for varying η_3 and $\eta_2 = t\eta_3$, $\eta_1 = t\eta_2$, at $t = 2, 3, 4$. **Right** (6-partition): $V_1(\eta_1, \eta_2, \eta_3, \eta_4, \eta_5)$ for varying η_5 and $\eta_i = t\eta_{i+1}$, at $t = 2, 3, 4$.

6.3 $\alpha = 2$ and $m = 5$

Numerically, the minimum of $V_2(\eta_1, \eta_2, \eta_3, \eta_4, \eta_5)$ is 2.106, attained at $\eta_1 = 0.893$, $\eta_2 = 0.339$, $\eta_3 = 0.184$, $\eta_4 = 0.111$, $\eta_5 = 0.068$. Figure 14 (right panel) plots $V_2(\eta_1, \eta_2, \eta_3, \eta_4, \eta_5)$ for varying η_5 and $\eta_i = t\eta_{i+1}$, $i = 4, 3, 2, 1$, as well as (left panel) $V_2(\eta_1, \eta_2, \eta_3)$ for varying η_3, and $\eta_2 = t\eta_3$, $\eta_1 = t\eta_2$.
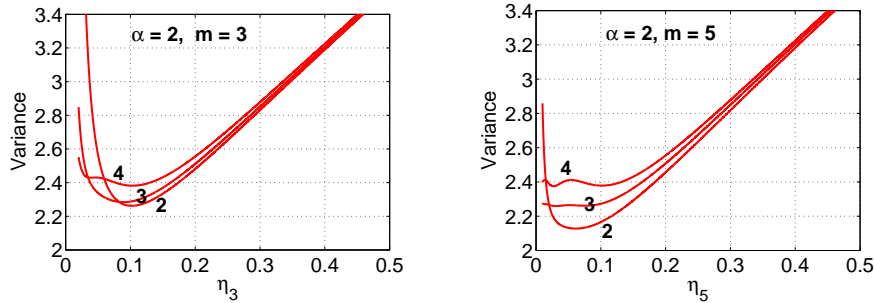

Figure 15: **Left** (4-partition): $V_2(\eta_1, \eta_2, \eta_3)$ for varying η_3 and $\eta_2 = t\eta_3$, $\eta_1 = t\eta_2$, at $t = 2, 3, 4$. **Right** (6-partition): $V_2(\eta_1, \eta_2, \eta_3, \eta_4, \eta_5)$ for varying η_5 and $\eta_i = t\eta_{i+1}$, at $t = 2, 3, 4$.

7 Extension and Future Work

Previously, [15] used counts and MLE, for improving classical minwise hashing and b-bit minwise hashing. In this paper, we focus on coding schemes for α-stable random projections on individual data vectors. We feel an important line of future work would be the study of coding schemes for analyzing the relation of two or multiple data vectors, which will be useful, for example, in the context of large-scale machine learning and efficient search/retrieval in massive data.

For example, [14] considered nonnegative data vectors under the sum-to-one constraint (i.e., the l_1 norm = 1). After applying Cauchy stable random projections separately on two data vectors, the collision probability of the two signs of the projected data is essentially monotonic in the χ^2 similarity (which is popular in computer vision). Now the open question is that, suppose we do not know the l_1 norms, how we should design coding schemes so that we can still evaluate the χ^2 similarity (or other similarities) using Cauchy random projections.

Another recent paper [13] re-visited classical Gaussian random projections (i.e., $\alpha = 2$). By assuming unit l_2 norms for the data vectors, [13] developed multi-bit coding schemes and estimators for the correlation between vectors. Can we, using just a few bits, still estimate the correlation if at the same time we must also estimate the l_2 norms?

8 Conclusion

Motivated by the recent work on "one scan 1-bit compressed sensing", we have developed 1-bit and multi-bit coding schemes for estimating the scale parameter of α-stable distributions. These simple coding schemes (even with just 1-bit) perform well in that, if the parameters are chosen appropriately, their variances are actually not much larger than the variances using full (i.e., infinite-bit) information. In general, using more bits increases the computational cost or storage cost (e.g., the cost of tabulations), with the benefits of stabilizing the performance so that the estimation variances do not increase much even when the parameters are far from optimal. In practice, we expect the $(m+1)$-partition scheme, combined with tabulation, for $m = 3$, 4, or 5, should be overall preferable. Here $m = 3$ corresponds to the 2-bit scheme, $m = 1$ to the 1-bit scheme.

A   Proof of Theorem 1 and Bias Corrections

The log-likelihood of the $n = n_1 + n_2$ observations is

$$l = n_1 \log F_\alpha \left(C/\Lambda_\alpha \right) + n_2 \log \left[1 - F_\alpha \left(C/\Lambda_\alpha \right) \right]$$

and its first derivative is

$$l' = \frac{\partial l}{\partial \Lambda_\alpha} = \left(-\frac{C}{\Lambda_\alpha^2} \right) \left(n_1 \frac{f_\alpha \left(C/\Lambda_\alpha \right)}{F_\alpha \left(C/\Lambda_\alpha \right)} + n_2 \frac{-f_\alpha \left(C/\Lambda_\alpha \right)}{1 - F_\alpha \left(C/\Lambda_\alpha \right)} \right)$$

$$= \left(-\frac{C}{\Lambda_\alpha^2} \right) \left(n_1 \frac{f}{F} - n_2 \frac{f}{1 - F} \right)$$

$$= \left(-\frac{C}{\Lambda_\alpha^2} \right) \left(\frac{n_1 - nF}{F - F^2} \right) f$$

For simplicity, we use F, f, f', f'' for $F_\alpha \left(C/\Lambda_\alpha \right)$, $f_\alpha \left(C/\Lambda_\alpha \right)$, $f'_\alpha \left(C/\Lambda_\alpha \right)$ and $f''_\alpha \left(C/\Lambda_\alpha \right)$, respectively. Setting $l' = 0$ leads to the MLE solution: $\frac{n_1}{n} = F_\alpha \left(C/\Lambda_\alpha \right)$, ie., $\hat{\Lambda}_\alpha = \frac{C}{F_\alpha^{-1}(n_1/n)}$.

According to classical statistical results [2, 20],

$$E \left(\hat{\Lambda}_\alpha \right) = \Lambda_\alpha - \frac{E(l')^3 + E(l'l'')}{2I^2} + O \left(\frac{1}{n^2} \right),$$

$$Var \left(\hat{\Lambda}_\alpha \right) = \frac{1}{I} + O \left(\frac{1}{n^2} \right)$$

where the Fisher Information $I = E(l')^2 = -E(l'')$. Here the derivatives l', l'', l''' are with respect to Λ_α. Thus, we need to computer the derivatives of l and evaluate their expectations.

By property of binomial distribution, we have $E(n_1) = nF_1$ and

$$E(n_1 - E(n_1)) = 0$$

$$E(n_1 - E(n_1))^2 = nF(1 - F)$$

$$E(n_1 - E(n_1))^3 = nF(1 - F)(1 - 2F)$$

Obviously

$$E(l') = \left(-\frac{C}{\Lambda_\alpha^2} \right) \left(\frac{nF - nF}{F - F^2} \right) f = 0$$

Furthermore,

$$E(l')^2 = \left(-\frac{C}{\Lambda_\alpha^2} \right)^2 f^2 \left(\frac{nF(1 - F)}{(F - F^2)^2} \right) = n \frac{C^2}{\Lambda_\alpha^4} \frac{f^2}{F(1 - F)} = I$$

$$E(l')^3 = \left(-\frac{C}{\Lambda_\alpha^2} \right)^3 f^3 \left(\frac{nF(1 - F)(1 - 2F)}{(F - F^2)^3} \right) = -n \frac{C^3}{\Lambda_\alpha^6} \frac{f^3(1 - 2F)}{F^2(1 - F)^2}$$

Next we work on the second derivative

$$l'' = -\frac{2}{\Lambda_\alpha} l' + \left(-\frac{C}{\Lambda_\alpha^2} \right)^2 \left(\frac{n_1 - nF}{F - F^2} \right) f' + \left(-\frac{C}{\Lambda_\alpha^2} \right)^2 f \left(\frac{-n_1 f + 2n_1 fF - nfF^2}{(F - F^2)^2} \right)$$

$$= -\frac{2}{\Lambda_\alpha} l' + \frac{f'}{f} \left(-\frac{C}{\Lambda_\alpha^2} \right) l' + \left(-\frac{C}{\Lambda_\alpha^2} \right) \frac{f}{1 - F} l' - \left(-\frac{C}{\Lambda_\alpha^2} \right)^2 n_1 \left(\frac{f^2}{F^2(1 - F)} \right)$$

$$E(l'') = \left(-\frac{C}{\Lambda_\alpha^2}\right)^2 f\left(\frac{-nfF + 2nfF^2 - nfF^2}{(F - F^2)^2}\right) = -n\frac{C^2}{\Lambda_\alpha^4}\frac{f^2}{F(1-F)} = -I$$

and higher-order derivatives

$$l''' = \left[-\frac{2}{\Lambda_\alpha}l' + \left(-\frac{C}{\Lambda_\alpha^2}\right)l'\left(\frac{f'}{f} + \frac{f}{1-F}\right) - \left(-\frac{C}{\Lambda_\alpha^2}\right)^2\left(\frac{n_1 f^2}{F^2(1-F)}\right)\right]'$$

$$= \frac{2}{\Lambda_\alpha^2}l' - \frac{2}{\Lambda_\alpha}l'' + \left(\frac{2C}{\Lambda_\alpha^3}\right)l'\left(\frac{f'}{f} + \frac{f}{1-F}\right) + \left(-\frac{C}{\Lambda_\alpha^2}\right)l''\left(\frac{f'}{f} + \frac{f}{1-F}\right)$$

$$+ \left(-\frac{C}{\Lambda_\alpha^2}\right)^2 l'\left(\frac{f''f - (f')^2}{f^2} + \frac{f'(1-F) + f^2}{(1-F)^2}\right)$$

$$+ \left(\frac{4C^2}{\Lambda_\alpha^5}\right)\left(\frac{n_1 f^2}{F^2(1-F)}\right) - \left(-\frac{C}{\Lambda_\alpha^2}\right)^3\left(\frac{2n_1 ff'(F^2 - F^3) - n_1 f^3(2F - 3F^2)}{F^4(1-F)^2}\right)$$

$$E(l'l'') = E\left[-\frac{2}{\Lambda_\alpha}(l')^2 + \left(-\frac{C}{\Lambda_\alpha^2}\right)(l')^2\left(\frac{f'}{f} + \frac{f}{1-F}\right) - \left(-\frac{C}{\Lambda_\alpha^2}\right)^2\left(\frac{f^2}{F^2(1-F)}\right)n_1 l'\right]$$

$$E(n_1 l') = \left(-\frac{C}{\Lambda_\alpha^2}\right)\left(\frac{E(n_1(n_1 - nF))}{F - F^2}\right)f = \left(-\frac{C}{\Lambda_\alpha^2}\right)\left(\frac{nF(1-F)}{F - F^2}\right)f = \left(-\frac{C}{\Lambda_\alpha^2}\right)nf$$

$$E(l'l'') = \left(-\frac{2}{\Lambda_\alpha} + \left(-\frac{C}{\Lambda_\alpha^2}\right)\left(\frac{f'}{f} + \frac{f}{1-F}\right)\right)n\frac{C^2}{\Lambda_\alpha^4}\frac{f^2}{F(1-F)} - \left(-\frac{C}{\Lambda_\alpha^2}\right)^3\left(\frac{nf^3}{F^2(1-F)}\right)$$

$$= \left(-\frac{2}{\Lambda_\alpha} + \left(-\frac{C}{\Lambda_\alpha^2}\right)\left(\frac{f'}{f}\right)\right)n\frac{C^2}{\Lambda_\alpha^4}\frac{f^2}{F(1-F)} - \left(-\frac{C}{\Lambda_\alpha^2}\right)^3\left(\frac{nf^3(1-2F)}{F^2(1-F)^2}\right)$$

$$E(l')^3 + E(l'l'') = \left(-\frac{2}{\Lambda_\alpha} + \left(-\frac{C}{\Lambda_\alpha^2}\right)\left(\frac{f'}{f}\right)\right)n\frac{C^2}{\Lambda_\alpha^4}\frac{f^2}{F(1-F)} = \left(-\frac{2}{\Lambda_\alpha} + \left(-\frac{C}{\Lambda_\alpha^2}\right)\left(\frac{f'}{f}\right)\right)I$$

Therefor, the bias-correction term is

$$\frac{E(l')^3 + E(l'l'')}{I^2} = \frac{\left(-\frac{2}{\Lambda_\alpha} + \left(-\frac{C}{\Lambda_\alpha^2}\right)\left(\frac{f'}{f}\right)\right)}{n\frac{C^2}{\Lambda_\alpha^4}\frac{f^2}{F(1-F)}} = -\frac{\Lambda_\alpha}{n}\frac{\left(2 + \frac{C}{\Lambda_\alpha}\frac{f'}{f}\right)}{\frac{C^2}{\Lambda_\alpha^2}\frac{f^2}{F(1-F)}} = -\frac{\Lambda_\alpha}{n}\frac{\left(2 + z\frac{f'}{f}\right)}{z^2\frac{f^2}{F(1-F)}}$$

$$= -\frac{\Lambda_\alpha}{n}\frac{n_1}{n}\left(1 - \frac{n_1}{n}\right)\frac{2 + z\frac{f'}{f}}{z^2 f^2}$$

where we denote $z = \frac{1}{\eta} = \frac{C}{\Lambda_\alpha}$. Note that since $l' = 0$, we have $z = \frac{C}{\Lambda_\alpha} = F_\alpha^{-1}(n_1/n)$, $F_\alpha(z) = \frac{n_1}{n}$.

Next, we derive more explicit expressions for $\alpha = 0+$, $\alpha = 1$, and $\alpha = 2$.

When $\alpha \to 0+$,

$$F_{0+}(z) = e^{-1/z}, \ f_{0+}(z) = \frac{1}{z^2}e^{-1/z}, \ f'_{0+}(z) = \frac{-2}{z^3}e^{-1/z} + \frac{1}{z^4}e^{-1/z}, \ F_{0+}^{-1}(z) = \frac{1}{\log 1/z}$$

20

$$2 + z\frac{f'(z)}{f(z)} = \frac{1}{z} = \frac{1}{F_\alpha^{-1}(n_1/n)} = \log n/n_1$$

$$z^2 f^2(z) = \frac{1}{z^2}e^{-2/z} = \frac{1}{z^2}e^{-2/F_\alpha^{-1}(n_1/n)} = \frac{\log^2(n/n_1)}{(n/n_1)^2}$$

$$\frac{E(l')^3 + E(l'l'')}{I^2} = -\frac{\Lambda_\alpha}{n}\frac{n_1}{n}\left(1 - \frac{n_1}{n}\right)\frac{2 + z\frac{f'}{f}}{z^2 f^2} = -\frac{\Lambda_\alpha}{n}\frac{n_1}{n}\left(1 - \frac{n_1}{n}\right)\frac{(n/n_1)^2}{\log n/n_1} = -\frac{\Lambda_\alpha}{n}\frac{(n/n_1 - 1)}{\log n/n_1}$$

Recall that $l' = 0 \Rightarrow \frac{n_1}{n} = F_\alpha(C/\Lambda_\alpha) \Rightarrow \frac{C}{\Lambda_\alpha} = F_\alpha^{-1}(n_1/n) = \frac{1}{\log n/n_1} \Rightarrow e^{\Lambda_\alpha/C} = n/n_1$.
Therefore, the bias-corrected MLE for $\alpha \to 0+$ is

$$\hat{\Lambda}_{0+,c} = \frac{C \log n/n_1}{1 + \frac{1/n_1 - 1/n}{2 \log n/n_1}}$$

When $\alpha = 1$, by properties of Cauchy distribution, we know

$$F_1(z) = \frac{2}{\pi}\tan^{-1} z, \quad f_1(z) = \frac{2}{\pi}\frac{1}{1 + z^2}, \quad f_1'(z) = \frac{2}{\pi}\frac{-2z}{(1 + z^2)^2}, \quad F_1^{-1}(z) = \tan\frac{\pi}{2}z$$

$$2 + z\frac{f'(z)}{f(z)} = 2 + \frac{-2z^2}{1 + z^2} = \frac{2}{1 + z^2}, \qquad z^2 f^2(z) = \frac{4}{\pi^2}\frac{z^2}{(1 + z^2)^2}$$

Note that $z = \frac{C}{\Lambda_\alpha} = F_\alpha^{-1}(n_1/n) = \tan\frac{\pi}{2}\frac{n_1}{n}$, $\tan^{-1} z = \frac{\pi}{2}\frac{n_1}{n}$. We have

$$\frac{2 + z\frac{f'}{f}}{z^2 f^2} = \frac{\pi^2}{2}\frac{1 + z^2}{z^2} = \frac{\pi^2}{2}\left(1 + \frac{1}{z^2}\right) = \frac{\pi^2}{2}\left(1 + \frac{1}{\tan^2 \frac{\pi}{2}\frac{n_1}{n}}\right)$$

$$\frac{E(l')^3 + E(l'l'')}{I^2} = -\frac{\Lambda_\alpha}{n}\frac{n_1}{n}\left(1 - \frac{n_1}{n}\right)\frac{2 + z\frac{f'}{f}}{z^2 f^2} = -\frac{\Lambda_\alpha}{n}\frac{n_1}{n}\left(1 - \frac{n_1}{n}\right)\frac{\pi^2}{2}\left(1 + \frac{1}{\tan^2 \frac{\pi}{2}\frac{n_1}{n}}\right)$$

Therefore, the bias-corrected MLE for $\alpha = 1$ is

$$\hat{\Lambda}_{1,c} = \frac{\frac{C}{\tan\frac{\pi}{2}\frac{n_1}{n}}}{1 + \frac{1}{n}\frac{\pi^2}{4}\frac{n_1}{n}\left(1 - \frac{n_1}{n}\right)\left(1 + \frac{1}{\tan^2 \frac{\pi}{2}\frac{n_1}{n}}\right)}$$

When $\alpha = 2$, since $S(2,1) \sim \sqrt{2} \times N(0,1)$, i.e., $|s_\alpha|^2 \sim 2\chi_1^2$, we have

$$F_2(z) = F_{\chi_1^2}(z/2) = 2\Phi(\sqrt{z/2}) - 1,$$

$$F_2^{-1}(t) = 2F_{\chi_1^2}^{-1}(t) = 2\left[\Phi^{-1}\left(\frac{t+1}{2}\right)\right]^2$$

$$z = F_2^{-1}(n_1/n) = 2\left[\Phi^{-1}\left(\frac{n_1/n + 1}{2}\right)\right]^2$$

$$f_2(z) = f_{\chi_1^2}(z/2)/2 = \frac{1}{2\sqrt{2\pi}}\frac{1}{\sqrt{z/2}}e^{-z/4} = \frac{1}{2\sqrt{\pi z}}e^{-z/4}$$

$$f_2'(z) = -\frac{1}{4\sqrt{\pi}z^{3/2}}e^{-z/4} - \frac{1}{8\sqrt{\pi}z^{1/2}}e^{-z/4}$$

$$z\frac{f_2'(z)}{f_2(z)} = -\frac{1}{2} - \frac{z}{4}, \qquad z^2 f^2(z) = \frac{z}{4\pi}e^{-z/2}$$

21

$$\frac{E(l')^3 + E(l'l'')}{I^2} = -\frac{\Lambda_\alpha}{n}\frac{n_1}{n}\left(1 - \frac{n_1}{n}\right)\frac{2 + z\frac{f'}{f}}{z^2 f^2} = -\frac{\Lambda_\alpha}{n}\frac{n_1}{n}\left(1 - \frac{n_1}{n}\right)\pi e^{z/2}\left(\frac{6}{z} - 1\right)$$

Therefore, the bias-corrected MLE for $\alpha = 2$ is

$$\hat{\Lambda}_{2,c} = \frac{\frac{C}{2F^{-1}_{\chi_1^2}(n_1/n)}}{1 + \frac{\pi}{n}\frac{n_1}{n}\left(1 - \frac{n_1}{n}\right)\left(\frac{3}{F^{-1}_{\chi_1^2}(n_1/n)} - 1\right)e^{\frac{F^{-1}_{\chi_1^2}(n_1/n)}{}}}$$

B Proof of Theorem 2

The task is to prove the following two bounds:

$$\mathbf{Pr}\left(\hat{\Lambda}_\alpha \geq (1 + \epsilon)\Lambda_\alpha\right) \leq \exp\left(-n\frac{\epsilon^2}{G_{R,\alpha,C,\epsilon}}\right), \ \epsilon \geq 0$$

$$\mathbf{Pr}\left(\hat{\Lambda}_\alpha \leq (1 - \epsilon)\Lambda_\alpha\right) \leq \exp\left(-n\frac{\epsilon^2}{G_{L,\alpha,C,\epsilon}}\right), \ 0 \leq \epsilon \leq 1$$

The proof is based on the expression of the MLE estimator $\hat{\Lambda}_\alpha = C/F_\alpha^{-1}(n_1/n)$, the fact that $n_1 \sim Binomial(n, F_\alpha(1/\eta))$, and Chernoff's original tail bounds [6] for the binomial distribution.

For the right tail bound, we have

$$\mathbf{Pr}\left(\hat{\Lambda}_\alpha \geq (1 + \epsilon)\Lambda_\alpha\right)$$

$$=\mathbf{Pr}\left(\frac{C}{F_\alpha^{-1}(n_1/n)} \geq (1 + \epsilon)\Lambda_\alpha\right)$$

$$=\mathbf{Pr}\left(\frac{n_1}{n} \leq F_\alpha\left(\frac{C}{(1 + \epsilon)\Lambda_\alpha}\right)\right)$$

$$=\mathbf{Pr}\left(\frac{n_1}{n} \leq F_\alpha\left(\frac{1}{(1 + \epsilon)\eta}\right)\right)$$

$$\leq \left[\frac{F_\alpha(1/\eta)}{F_\alpha(1/(1 + \epsilon)\eta)}\right]^{nF_\alpha(1/(1+\epsilon)\eta)} \times \left[\frac{1 - F_\alpha(1/\eta)}{1 - F_\alpha(1/(1 + \epsilon)\eta)}\right]^{n-nF_\alpha(1/(1+\epsilon)\eta)}$$

$$=\exp\left(-n\frac{\epsilon^2}{G_{R,\alpha,C,\epsilon}}\right)$$

where

$$\frac{\epsilon^2}{G_{R,\alpha,C,\epsilon}} = -F_\alpha(1/(1 + \epsilon)\eta)\log\left[\frac{F_\alpha(1/\eta)}{F_\alpha(1/(1 + \epsilon)\eta)}\right] - (1 - F_\alpha(1/(1 + \epsilon)\eta))\log\left[\frac{1 - F_\alpha(1/\eta)}{1 - F_\alpha(1/(1 + \epsilon)\eta)}\right]$$

Next, for the left tail bound, we have

$$
\begin{aligned}
&\mathbf{Pr}\left(\hat{\Lambda}_\alpha \le (1-\epsilon)\Lambda_\alpha\right)\\
={}&\mathbf{Pr}\left(\frac{C}{F_\alpha^{-1}(n_1/n)} \le (1-\epsilon)\Lambda_\alpha\right)\\
={}&\mathbf{Pr}\left(\frac{n_1}{n} \ge F_\alpha\left(\frac{C}{(1-\epsilon)\Lambda_\alpha}\right)\right)\\
={}&\mathbf{Pr}\left(\frac{n_1}{n} \ge F_\alpha\left(\frac{1}{(1-\epsilon)\eta}\right)\right)\\
\le{}&\left[\frac{F_\alpha(1/\eta)}{F_\alpha(1/(1-\epsilon)\eta)}\right]^{nF_\alpha(1/(1-\epsilon)\eta)} \times \left[\frac{1-F_\alpha(1/\eta)}{1-F_\alpha(1/(1-\epsilon)\eta)}\right]^{n-nF_\alpha(1/(1-\epsilon)\eta)}\\
={}&\exp\left(-n\frac{\epsilon^2}{G_{L,\alpha,C,\epsilon}}\right)
\end{aligned}
$$

where

$$
\frac{\epsilon^2}{G_{L,\alpha,C,\epsilon}} = -F_\alpha(1/(1-\epsilon)\eta)\log\left[\frac{F_\alpha(1/\eta)}{F_\alpha(1/(1-\epsilon)\eta)}\right] - (1-F_\alpha(1/(1-\epsilon)\eta))\log\left[\frac{1-F_\alpha(1/\eta)}{1-F_\alpha(1/(1-\epsilon)\eta)}\right]
$$

C   Proof of Theorem 3

With the 2-bit scheme, we need to introduce 3 threshold values: $C_1 \le C_2 \le C_3$, and define

$$
\begin{aligned}
p_1 &= \mathbf{Pr}\left(z_j \le C_1\right) = F_\alpha\left(C_1/\Lambda_\alpha\right)\\
p_2 &= \mathbf{Pr}\left(C_1 < z_j \le C_2\right) = F_\alpha\left(C_2/\Lambda_\alpha\right) - F_\alpha\left(C_1/\Lambda_\alpha\right)\\
p_3 &= \mathbf{Pr}\left(C_2 < z_j \le C_3\right) = F_\alpha\left(C_3/\Lambda_\alpha\right) - F_\alpha\left(C_2/\Lambda_\alpha\right)\\
p_4 &= \mathbf{Pr}\left(z_j > C_3\right) = 1 - F_\alpha\left(C_3/\Lambda_\alpha\right)
\end{aligned}
$$

and

$$
\begin{aligned}
n_1 &= \sum_{j=1}^{n} 1\{z_j \le C_1\}, & n_2 &= \sum_{j=1}^{n} 1\{C_1 < z_j \le C_2\}\\
n_3 &= \sum_{j=1}^{n} 1\{C_2 < z_j \le C_3\}, & n_4 &= \sum_{j=1}^{n} 1\{z_j > C_3\}
\end{aligned}
$$

The log-likelihood of these $n = n_1 + n_2 + n_3 + n_4$ observations can be expressed as

$$
\begin{aligned}
l ={}& n_1 \log p_1 + n_2 \log p_2 + n_3 \log p_3 + n_4 \log p_4\\
={}& n_1 \log F_\alpha\left(C_1/\Lambda_\alpha\right) + n_2 \log\left[F_\alpha\left(C_2/\Lambda_\alpha\right) - F_\alpha\left(C_1/\Lambda_\alpha\right)\right]\\
&+ n_3 \log\left[F_\alpha\left(C_3/\Lambda_\alpha\right) - F_\alpha\left(C_2/\Lambda_\alpha\right)\right] + n_4 \log\left[1 - F_\alpha\left(C_3/\Lambda_\alpha\right)\right]\\
={}& n_1 \log F_1 + n_2 \log(F_2 - F_1) + n_3 \log(F_3 - F_2) + n_4 \log(1 - F_3)
\end{aligned}
$$

To seek the MLE of Λ_α, we need to compute the first derivative:

$$l' = \frac{\partial l}{\partial \Lambda_\alpha} = n_1 \frac{f_\alpha\left(C_1/\Lambda_\alpha\right)}{F_\alpha\left(C_1/\Lambda_\alpha\right)} \left(-\frac{C_1}{\Lambda_\alpha^2}\right) + n_2 \frac{f_\alpha\left(C_2/\Lambda_\alpha\right)\left(-\frac{C_2}{\Lambda_\alpha^2}\right) - f_\alpha\left(C_1/\Lambda_\alpha\right)\left(-\frac{C_1}{\Lambda_\alpha^2}\right)}{F_\alpha\left(C_2/\Lambda_\alpha\right) - F_\alpha\left(C_1/\Lambda_\alpha\right)}$$

$$+ n_3 \frac{f_\alpha\left(C_3/\Lambda_\alpha\right)\left(-\frac{C_3}{\Lambda_\alpha^2}\right) - f_\alpha\left(C_2/\Lambda_\alpha\right)\left(-\frac{C_2}{\Lambda_\alpha^2}\right)}{F_\alpha\left(C_3/\Lambda_\alpha\right) - F_\alpha\left(C_2/\Lambda_\alpha\right)} + n_4 \frac{-f_\alpha\left(C_3/\Lambda_\alpha\right)}{1 - F_\alpha\left(C_3/\Lambda_\alpha\right)}\left(-\frac{C_3}{\Lambda_\alpha^2}\right)$$

$$= n_1 \frac{\left(-\frac{C_1}{\Lambda_\alpha^2}\right)f_1}{F_1} + n_2 \frac{\left(-\frac{C_2}{\Lambda_\alpha^2}\right)f_2 - \left(-\frac{C_1}{\Lambda_\alpha^2}\right)f_1}{F_2 - F_1} + n_3 \frac{\left(-\frac{C_3}{\Lambda_\alpha^2}\right)f_3 - \left(-\frac{C_2}{\Lambda_\alpha^2}\right)f_2}{F_3 - F_2} + n_4 \frac{-\left(-\frac{C_3}{\Lambda_\alpha^2}\right)f_3}{1 - F_3}$$

Since $E(n_1) = nF_1$, $E(n_2) = n(F_2 - F_1)$, $E(n_3) = n(F_3 - F_2)$, $E(n_4) = n(1 - F_3)$, we have

$$E(l') = \left(-\frac{C_1}{\Lambda_\alpha^2}\right)f_1 + \left(-\frac{C_2}{\Lambda_\alpha^2}\right)f_2 - \left(-\frac{C_1}{\Lambda_\alpha^2}\right)f_1 + \left(-\frac{C_3}{\Lambda_\alpha^2}\right)f_3 - \left(-\frac{C_2}{\Lambda_\alpha^2}\right)f_2 - \left(-\frac{C_3}{\Lambda_\alpha^2}\right)f_3 = 0$$

Next, we compute the Fisher Information,

$$l'' = n_1 \frac{\left(-\frac{C_1}{\Lambda_\alpha^2}\right)^2 f_1' F_1 - \left(-\frac{C_1}{\Lambda_\alpha^2}\right)^2 (f_1)^2}{F_1^2}$$

$$+ n_2 \frac{\left[\left(-\frac{C_2}{\Lambda_\alpha^2}\right)^2 f_2' - \left(-\frac{C_1}{\Lambda_\alpha^2}\right)^2 f_1'\right](F_2 - F_1) - \left[\left(-\frac{C_2}{\Lambda_\alpha^2}\right)f_2 - \left(-\frac{C_1}{\Lambda_\alpha^2}\right)f_1\right]^2}{(F_2 - F_1)^2}$$

$$+ n_3 \frac{\left[\left(-\frac{C_3}{\Lambda_\alpha^2}\right)^2 f_3' - \left(-\frac{C_2}{\Lambda_\alpha^2}\right)^2 f_2'\right](F_3 - F_2) - \left[\left(-\frac{C_3}{\Lambda_\alpha^2}\right)f_3 - \left(-\frac{C_2}{\Lambda_\alpha^2}\right)f_2\right]^2}{(F_3 - F_2)^2}$$

$$+ n_4 \frac{\left[-\left(-\frac{C_3}{\Lambda_\alpha^2}\right)^2 f_3'\right](1 - F_3) - \left[-\left(-\frac{C_3}{\Lambda_\alpha^2}\right)f_3\right]^2}{(1 - F_3)^2}$$

$$-I = E(l'') = \frac{\left(-\frac{C_1}{\Lambda_\alpha^2}\right)^2 f_1' F_1 - \left(-\frac{C_1}{\Lambda_\alpha^2}\right)^2 (f_1)^2}{F_1}$$

$$+ \frac{\left[\left(-\frac{C_2}{\Lambda_\alpha^2}\right)^2 f_2' - \left(-\frac{C_1}{\Lambda_\alpha^2}\right)^2 f_1'\right](F_2 - F_1) - \left[\left(-\frac{C_2}{\Lambda_\alpha^2}\right)f_2 - \left(-\frac{C_1}{\Lambda_\alpha^2}\right)f_1\right]^2}{(F_2 - F_1)}$$

$$+ \frac{\left[\left(-\frac{C_3}{\Lambda_\alpha^2}\right)^2 f_3' - \left(-\frac{C_2}{\Lambda_\alpha^2}\right)^2 f_2'\right](F_3 - F_2) - \left[\left(-\frac{C_3}{\Lambda_\alpha^2}\right)f_3 - \left(-\frac{C_2}{\Lambda_\alpha^2}\right)f_2\right]^2}{(F_3 - F_2)}$$

$$+ \frac{\left[-\left(-\frac{C_3}{\Lambda_\alpha^2}\right)^2 f_3'\right](1 - F_3) - \left[-\left(-\frac{C_3}{\Lambda_\alpha^2}\right)f_3\right]^2}{(1 - F_3)}$$

$$= -\frac{\left(-\frac{C_1}{\Lambda_\alpha^2}\right)^2 (f_1)^2}{F_1} - \frac{\left[\left(-\frac{C_2}{\Lambda_\alpha^2}\right)f_2 - \left(-\frac{C_1}{\Lambda_\alpha^2}\right)f_1\right]^2}{(F_2 - F_1)} - \frac{\left[\left(-\frac{C_3}{\Lambda_\alpha^2}\right)f_3 - \left(-\frac{C_2}{\Lambda_\alpha^2}\right)f_2\right]^2}{(F_3 - F_2)} - \frac{\left[-\left(-\frac{C_3}{\Lambda_\alpha^2}\right)f_3\right]^2}{(1 - F_3)}$$

The asymptotic bias is

$$E\left(\hat{\Lambda}_\alpha\right) = \Lambda_\alpha - \frac{E(l')^3 + E(l'l'')}{2I^2} + O\left(\frac{1}{n^2}\right)$$

For convenience, we re-write l' and l'' as follows.

$$l' = [n_1 - nF_1] \frac{\left(-\frac{C_1}{\Lambda_\alpha^2}\right) f_1}{F_1} + [n_2 - n(F_2 - F_1)] \frac{\left(-\frac{C_2}{\Lambda_\alpha^2}\right) f_2 - \left(-\frac{C_1}{\Lambda_\alpha^2}\right) f_1}{F_2 - F_1}$$

$$+ [n_3 - n(F_3 - F_2)] \frac{\left(-\frac{C_3}{\Lambda_\alpha^2}\right) f_3 - \left(-\frac{C_2}{\Lambda_\alpha^2}\right) f_2}{F_3 - F_2} + [n_4 - n(1 - F_3)] \frac{-\left(-\frac{C_3}{\Lambda_\alpha^2}\right) f_3}{1 - F_3}$$

$$= \sum_{i=1}^4 z_i p_i'/p_i, \quad \text{where} \quad z_i = n_i - np_i, \quad p_i' = \frac{\partial p_i}{\partial \Lambda_\alpha}$$

$$l'' = [n_1 - nF_1] \frac{\left(-\frac{C_1}{\Lambda_\alpha^2}\right)^2 f_1' F_1 - \left(-\frac{C_1}{\Lambda_\alpha^2}\right)^2 (f_1)^2}{F_1^2}$$

$$+ [n_2 - n(F_2 - F_1)] \frac{\left[\left(-\frac{C_2}{\Lambda_\alpha^2}\right)^2 f_2' - \left(-\frac{C_1}{\Lambda_\alpha^2}\right)^2 f_1'\right](F_2 - F_1) - \left[\left(-\frac{C_2}{\Lambda_\alpha^2}\right) f_2 - \left(-\frac{C_1}{\Lambda_\alpha^2}\right) f_1\right]^2}{(F_2 - F_1)^2}$$

$$+ [n_3 - n(F_3 - F_2)] \frac{\left[\left(-\frac{C_3}{\Lambda_\alpha^2}\right)^2 f_3' - \left(-\frac{C_2}{\Lambda_\alpha^2}\right)^2 f_2'\right](F_3 - F_2) - \left[\left(-\frac{C_3}{\Lambda_\alpha^2}\right) f_3 - \left(-\frac{C_2}{\Lambda_\alpha^2}\right) f_2\right]^2}{(F_3 - F_2)^2}$$

$$+ [n_4 - n(1 - F_3)] \frac{\left[-\left(-\frac{C_3}{\Lambda_\alpha^2}\right)^2 f_3'\right](1 - F_3) - \left[-\left(-\frac{C_3}{\Lambda_\alpha^2}\right) f_3\right]^2}{(1 - F_3)^2}$$

$$- n\frac{\left(-\frac{C_1}{\Lambda_\alpha^2}\right)^2 (f_1)^2}{F_1} - n\frac{\left[\left(-\frac{C_2}{\Lambda_\alpha^2}\right) f_2 - \left(-\frac{C_1}{\Lambda_\alpha^2}\right) f_1\right]^2}{(F_2 - F_1)} - n\frac{\left[\left(-\frac{C_3}{\Lambda_\alpha^2}\right) f_3 - \left(-\frac{C_2}{\Lambda_\alpha^2}\right) f_2\right]^2}{(F_3 - F_2)} - n\frac{\left[-\left(-\frac{C_3}{\Lambda_\alpha^2}\right) f_3\right]^2}{(1 - F_3)}$$

$$= \sum_{i=1}^4 z_i \frac{p_i'' p_i - (p_i')^2}{p_i^2} - I, \quad \text{where} \quad p_i'' = \frac{\partial^2 p_i}{\partial \Lambda_\alpha^2}$$

We will take advantage of the central comments of multinomial:

$$E((n_i - np_i)^2) = np_i(1 - p_i)$$
$$E((n_i - np_i)(n_j - np_j)) = -np_ip_j \quad (i \neq j)$$
$$E((n_i - np_i)^3) = np_i(1 - p_i)(1 - 2p_i)$$
$$E((n_i - np_i)^2(n_j - np_j)) = -np_ip_j(1 - 2p_i) \quad (i \neq j)$$
$$E((n_i - np_i)(n_j - np_j)(n_k - np_k)) = 2np_ip_jp_k \quad (i \neq j \neq k)$$

and the following expansion,

$$(a + b + c + d)^2 = a^3 + b^3 + c^3 + d^3 + 3a^2b + 3a^2c + 3a^2d + 3ab^2 + 3b^2c + 3b^2d + 3ac^2 + 3bc^2$$
$$+ 3c^2d + 3ad^2 + 3bd^2 + 3cd^2 + 6abc + 6abd + 6acd + 6bcd$$

We are now ready to compute $E(l')^3$. Because

$$l' = \sum_{i=1}^{4} z_i p'_i / p_i, \qquad \text{where} \;\; z_i = n_i - np_i, \;\; p'_i = \frac{\partial p_i}{\partial \Lambda_\alpha}$$

we need to compute

$$E\left(\left(z_1 \frac{p'_1}{p_1} \right)^2 \left(z_2 \frac{p'_2}{p_2} + z_3 \frac{p'_3}{p_3} + z_4 \frac{p'_4}{p_4} \right) \right)$$

$$= -np_1(1 - 2p_1)\frac{(p'_1)^2}{p_1^2} \left(p_2 \frac{p'_2}{p_2} + p_3 \frac{p'_3}{p_3} + p_4 \frac{p'_4}{p_4} \right)$$

$$= -n(1 - 2p_1)\frac{(p'_1)^2}{p_1} \left(p'_2 + p'_3 + p'_4 \right)$$

$$= n(1 - 2p_1)\frac{(p'_1)^3}{p_1}$$

and

$$E\left(z_1 \frac{p'_1}{p_1} z_2 \frac{p'_2}{p_2} z_3 \frac{p'_3}{p_3} \right) = 2np'_1 p'_2 p'_3$$

Thus

$$E(l')^3 = \sum_{i=1}^{4} n(1 - p_i)(1 - 2p_i)\frac{(p'_i)^3}{p_i^2} + 3\sum_{i=1}^{4} n(1 - 2p_i)\frac{(p'_i)^3}{p_i}$$

$$+ 12np'_1 p'_2 p'_3 + 12np'_1 p'_2 p'_4 + 12np'_2 p'_3 p'_4 + 12np'_1 p'_3 p'_4$$

$$= n\sum_{i=1}^{4} \frac{(p'_i)^3}{p_i^2} - 4n\sum_{i=1}^{4} (p'_i)^3 + 12np'_1 p'_2 p'_3 + 12np'_1 p'_2 p'_4 + 12np'_2 p'_3 p'_4 + 12np'_1 p'_3 p'_4$$

Next, we compute $E\left(l'l''\right)$.

$$E\left(l'l''\right) = E\left[\left(\sum_{i=1}^{4} z_i \frac{p'_i}{p_i} \right) \left(\sum_{i=1}^{4} z_i \frac{p''_i p_i - (p'_i)^2}{p_i^2} \right) \right]$$

$$= E\left[\sum_{i=1}^{4} z_i^2 \frac{p''_i p'_i p_i - (p'_i)^3}{p_i^3} + \sum_{i \neq j} z_i z_j \frac{p'_i}{p_i} \frac{p''_j p_j - (p'_j)^2}{p_j^2} \right]$$

$$= n\sum_{i=1}^{4} (1 - p_i)\frac{p''_i p'_i p_i - (p'_i)^3}{p_i^2} - n\sum_{i \neq j} p'_i \frac{p''_j p_j - (p'_j)^2}{p_j}$$

$$= n\sum_{i=1}^{4} \frac{p''_i p'_i p_i - (p'_i)^3}{p_i^2} - n\sum_{i=1}^{4} p''_i p'_i + n\sum_{i=1}^{4} \frac{(p'_i)^3}{p_i} - n\sum_{i \neq j} p'_i p''_j + n\sum_{i \neq j} \frac{p'_i (p'_j)^2}{p_j}$$

$$= n\sum_{i=1}^{4} \frac{p''_i p'_i}{p_i} - n\sum_{i=1}^{4} \frac{(p'_i)^3}{p_i^2} + n\left(\sum_{i=1}^{4} p'_i \right)\left(\sum_{i=1}^{4} \frac{(p'_i)^2}{p_i} \right) - n\left(\sum_{i=1}^{4} p'_i \right)\left(\sum_{i=1}^{4} p''_i \right)$$

$$= n\sum_{i=1}^{4} \frac{p''_i p'_i}{p_i} - n\sum_{i=1}^{4} \frac{(p'_i)^3}{p_i^2}$$

$$E\left(l'l''\right)+E\left(l'\right)^3 = n\sum_{i=1}^{4}\frac{p_i''p_i'}{p_i} - 4n\sum_{i=1}^{4}(p_i')^3 + 12np_1'p_2'p_3' + 12np_1'p_2'p_4' + 12np_2'p_3'p_4' + 12np_1'p_3'p_4'$$

$$= n\sum_{i=1}^{4}\frac{p_i''p_i'}{p_i}$$

To see this, we can use the fact that $1 = \sum_{i=1}^{4}p_i$, $0 = \sum_{i=1}^{4}p_i'$, and

$$0 = \left(\sum_{i=1}^{4}p_i'\right)^3 = \sum_{i=1}^{4}(p_i')^3 + 3\sum_{i=1}^{4}(p_i')^2(-p_i') + 6p_1'p_2'p_3' + 6p_1'p_2'p_4' + 6p_2'p_3'p_4' + 6p_1'p_3'p_4'$$

$$= -2\sum_{i=1}^{4}(p_i')^3 + 6p_1'p_2'p_3' + 6p_1'p_2'p_4' + 6p_2'p_3'p_4' + 6p_1'p_3'p_4'$$

Therefore,

$$E\left(l'l''\right)+E\left(l'\right)^3 = n\sum_{i=1}^{4}\frac{p_i''p_i'}{p_i}$$

$$= n\frac{\left(-\frac{C_1}{\Lambda_\alpha^2}\right)f_1\left[2\frac{C_1}{\Lambda_\alpha^3}f_1 + \left(-\frac{C_1}{\Lambda_\alpha^2}\right)^2 f_1'\right]}{F_1}$$

$$+ n\frac{\left[\left(-\frac{C_2}{\Lambda_\alpha^2}\right)f_2 - \left(-\frac{C_1}{\Lambda_\alpha^2}\right)f_1\right]\left[2\frac{C_2}{\Lambda_\alpha^3}f_2 + \left(-\frac{C_2}{\Lambda_\alpha^2}\right)^2 f_2' - 2\frac{C_1}{\Lambda_\alpha^3}f_1 - \left(-\frac{C_1}{\Lambda_\alpha^2}\right)^2 f_1'\right]}{F_2 - F_1}$$

$$+ n\frac{\left[\left(-\frac{C_3}{\Lambda_\alpha^2}\right)f_3 - \left(-\frac{C_2}{\Lambda_\alpha^2}\right)f_2\right]\left[2\frac{C_3}{\Lambda_\alpha^3}f_3 + \left(-\frac{C_3}{\Lambda_\alpha^2}\right)^2 f_3' - 2\frac{C_2}{\Lambda_\alpha^3}f_2 - \left(-\frac{C_2}{\Lambda_\alpha^2}\right)^2 f_2'\right]}{F_3 - F_2}$$

$$+ n\frac{\left(-\frac{C_3}{\Lambda_\alpha^2}\right)f_3\left[2\frac{C_3}{\Lambda_\alpha^3}f_3 + \left(-\frac{C_3}{\Lambda_\alpha^2}\right)^2 f_3'\right]}{1 - F_3}$$

Because

$$I = n\frac{\left(-\frac{C_1}{\Lambda_\alpha^2}\right)^2 (f_1)^2}{F_1} + n\frac{\left[\left(-\frac{C_2}{\Lambda_\alpha^2}\right)f_2 - \left(-\frac{C_1}{\Lambda_\alpha^2}\right)f_1\right]^2}{(F_2 - F_1)} + n\frac{\left[\left(-\frac{C_3}{\Lambda_\alpha^2}\right)f_3 - \left(-\frac{C_2}{\Lambda_\alpha^2}\right)f_2\right]^2}{(F_3 - F_2)} + n\frac{\left[-\left(-\frac{C_3}{\Lambda_\alpha^2}\right)f_3\right]^2}{(1 - F_3)}$$

and

$$\frac{E\left(l'l''\right)+E\left(l'\right)^3}{2I^2} = \frac{\Lambda_\alpha}{n}\left(-\frac{1}{B} + \frac{D}{2B^2}\right)$$

we have

$$E\left(\hat{\Lambda}_\alpha\right) = \Lambda_\alpha - \frac{\Lambda_\alpha}{n}\left(-\frac{1}{B} + \frac{D}{2B^2}\right) + O\left(\frac{1}{n^2}\right) = \Lambda_\alpha\left(1 + \frac{1}{nB} - \frac{D}{2nB^2}\right) + O\left(\frac{1}{n^2}\right)$$

which leads to a bias-corrected estimator

$$\hat{\Lambda}_{\alpha,c} = \frac{\hat{\Lambda}_\alpha}{1 + \frac{1}{nB} - \frac{D}{2nB^2}}$$

where

$$B = \frac{\left(-\frac{C_1}{\Lambda_\alpha}\right)^2 f_1^2}{F_1} + \frac{\left[\left(-\frac{C_2}{\Lambda_\alpha}\right) f_2 - \left(-\frac{C_1}{\Lambda_\alpha}\right) f_1\right]^2}{F_2 - F_1} + \frac{\left[\left(-\frac{C_3}{\Lambda_\alpha}\right) f_3 - \left(-\frac{C_2}{\Lambda_\alpha}\right) f_2\right]^2}{F_3 - F_2} + \frac{\left(-\frac{C_3}{\Lambda_\alpha}\right)^2 f_3^2}{1 - F_3}$$

and

$$D = \frac{\left(-\frac{C_1}{\Lambda_\alpha}\right)^3 f_1 f_1'}{F_1} + \frac{\left[\left(-\frac{C_2}{\Lambda_\alpha}\right) f_2 - \left(-\frac{C_1}{\Lambda_\alpha}\right) f_1\right] \left[\left(-\frac{C_2}{\Lambda_\alpha}\right)^2 f_2' - \left(-\frac{C_1}{\Lambda_\alpha}\right)^2 f_1'\right]}{F_2 - F_1}$$

$$+ \frac{\left[\left(-\frac{C_3}{\Lambda_\alpha}\right) f_3 - \left(-\frac{C_2}{\Lambda_\alpha}\right) f_2\right] \left[\left(-\frac{C_3}{\Lambda_\alpha}\right)^2 f_3' - \left(-\frac{C_2}{\Lambda_\alpha}\right)^2 f_2'\right]}{F_3 - F_2} + \frac{\left(-\frac{C_3}{\Lambda_\alpha}\right)^3 f_3 f_3'}{1 - F_3}$$

References

[1] N. Alon, Y. Matias, and M. Szegedy. The space complexity of approximating the frequency moments. In *STOC*, pages 20–29, Philadelphia, PA, 1996.

[2] M. S. Bartlett. Approximate confidence intervals, II. *Biometrika*, 40(3/4):306–317, 1953.

[3] P. Boufounos and R. Baraniuk. 1-bit compressive sensing. In *Information Sciences and Systems, 2008.*, pages 16–21, March 2008.

[4] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, Feb 2006.

[5] J. M. Chambers, C. L. Mallows, and B. W. Stuck. A method for simulating stable random variables. *Journal of the American Statistical Association*, 71(354):340–344, 1976.

[6] H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics*, 23(4):493–507, 1952.

[7] N. Cressie. A note on the behaviour of the stable distributions for small index. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 31(1):61–64, 1975.

[8] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, April 2006.

[9] P. Indyk. Stable distributions, pseudorandom generators, embeddings, and data stream computation. *Journal of ACM*, 53(3):307–323, 2006.

[10] L. Jacques, J. N. Laska, P. T. Boufounos, and R. G. Baraniuk. Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors. *IEEE Transactions on Information Theory*, 59(4):2082–2102, 2013.

[11] P. Li. Estimators and tail bounds for dimension reduction in l_α $(0 < \alpha \le 2)$ using stable random projections. In *SODA*, pages 10 – 19, San Francisco, CA, 2008.

[12] P. Li. One scan 1-bit compressed sensing. Technical report, arXiv:1503.02346, 2015.

[13] P. Li, M. Mitzenmacher, and A. Shrivastava. Coding for random projections. In *ICML*, 2014.

[14] P. Li, G. Samorodnitsky, and J. Hopcroft. Sign cauchy projections and chi-square kernel. In *NIPS*, Lake Tahoe, NV, 2013.

[15] P. Li and A. C. König. Accurate estimators for improving minwise hashing and b-bit minwise hashing. Technical report, arXiv:1108.0895, 2011.

[16] P. Li, C.-H. Zhang, and T. Zhang. Compressed counting meets compressed sensing. In *COLT*, 2014.

[17] S. Muthukrishnan. Data streams: Algorithms and applications. *Foundations and Trends in Theoretical Computer Science*, 1:117–236, 2005.

[18] Y. Plan and R. Vershynin. Robust 1-bit compressed sensing and sparse logistic regression: A convex programming approach. *IEEE Transactions on Information Theory*, 59(1):482–494, 2013.

[19] G. Samorodnitsky and M. S. Taqqu. *Stable Non-Gaussian Random Processes*. Chapman & Hall, New York, 1994.

[20] L. R. Shenton and K. O. Bowman. Higher moments of a maximum-likelihood estimate. *Journal of Royal Statistical Society B*, 25(2):305–317, 1963.

[21] M. Slawski and P. Li. b-bit marginal regression. In *NIPS*, Montreal, CA, 2015.

[22] V. M. Zolotarev. *One-dimensional Stable Distributions*. American Mathematical Society, Providence, RI, 1986.